Neural Representation of Working Memory Contents At Different Levels of Abstraction

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum naturalium (Dr. rer. nat.)

eingereicht an der

Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

von

Chang Yan, M. Sc.

Präsidentin der Humboldt-Universität zu Berlin: Prof. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät: Prof. Dr. Bernhard Grimm

Gutachter/Gutachterin:

1. Prof. Dr. John-Dylan Haynes
2. Prof. Dr. Felix Blankenburg
3. Prof. Dr. Philipp Sterzer

Tag der mündlichen Prüfung: 31.08.2020

https://doi.org/10.18452/22232

To all the moments

of curiosity, concentration, and self-challenge

Table of Contents

Abbreviations

aBA	Anterior Broca's Area
ANOVA	Analysis of Variance
BCAN	Berlin Center for Advanced Neuroimaging
BF	Basis Function
BOLD	Blood-Oxygenation-Level-Dependent
CI	Confidence Interval
CI task	Category Identification task
CIE	Commission Internationale de l'Eclairage
CN task	Category Naming task
CV	Cross Validation
cvMANOVA	Cross-Validated Multivariate Analysis of Variance
D	(Pattern) Distinctness
DE task	Delayed Estimation task
dHb	Deoxygenated Haemoglobin
dlPFC	Dorsolateral Prefrontal Cortex
EPI	Echo-Planar Imaging
EVC	Early Visual Cortex
FIR	Finite Impulse Response
fMRI	Functional Magnetic Resonance Imaging
FWE	Family-Wise Error
FWHM	Full-Width at Half Maximum
GLM	General Linear Model
Hb	Oxygenated Hemoglobin
HRF	Hemodynamic Response Function
IEM	Inverted Encoding Model
IPS	Intraparietal Sulcus
ITI	Inter-Trial Interval
LDA	Linear Discriminant Analysis
lPFC	Lateral Prefrontal Cortex
lPMC	Left Premotor Cortex
LTM	Long-Term Memory

MNI	Montreal Neurological Institute
MPRAGE	Magnetization-Prepared Rapid Gradient Echo
MRI	Magnetic Resonance Imaging
MVPA	Multivariate Pattern Analysis
MGLM	Multivariate General Linear Model
NMR	Nuclei Magnetic Resonance
PFC	Prefrontal Cortex
RF	Radio Frequency
ROI	Region of Interest
SEM	Standard Error of Mean
SM	Sensory Memory
STM	Short-Term Memory
SVM	Support Vector Machine
TE	Echo Time
TR	Repetition Time
T1	Longitudinal Relaxation Time
T2*	Effective Transversal Relaxation Time
UDE task	Undelayed Estimation task
WM	Working Memory

Acknowledgement

The long and challenging journey of pursuing science is lonely, but I would have never completed this doctoral thesis without the help, support and encouragement from many people. Therefore, I want to acknowledge them for that.

First of all, I want to express my sincere gratitude to Prof. John-Dylan Haynes, who invited me to do science, and has given me many support, encouragement and guidance for years. Next, I want to express my big thankfulness to Dr. Thomas B. Christophel, with whom I had many enthusiastic discussions, for advising me extensively throughout my PHD. I also want to thank Thomas for always responding super-fast to my Emails.

I am grateful to our lab secretary: Brigitte, and BCCN coordinators: Vanessa, Robert, Margret and Lisa, who have given me much help. I also want to express my thankfulness to my coach Julia Lemmle, to Chris Donkin and Martina Michalikova, who have provided me with inspiration and suggestions. I also would like to thank my lovely office-mates: Polina, Shikhar, Jan, Felix W, Patrick, and my lab colleagues: Doris, Joram, Yi, Kai, Carsten A, Carsten B, Riccardo, Felix T, Corinna, Kerstin, Martin W, Fabian, Stefan Hetzer, Sebastian.

Many friends, with whom I studied together and had interesting long conversations, have given me enormous encouragement. I want to say a big thanks to Wendelin, Lei, Lena Li, Jeng, Xu Cheng, Yan Chen, Jiani, Weijie, Xingxing, Betty, Colleen, DJ, Paul, Arseny, Shuyan, Shulin Gao, Yuchen, Chanjuan, Junyu, Yang Zhang, Yang Ni, Yefei Yin, Iila Li, Athena Chen, Xuemei and Yang Xin.

Finally and especially, I would like to thank my mum, my father, my grandpa (who was so proud of me and would be so happy to see me becoming a ‘Dr. rer. nat.’), my sister Qiao, and my other family members.

Above is merely an incomplete list, there are many more people I am grateful for, but time is not on my side when I write these lines.

Abstract

Research on the neural basis of working memory (WM) has received broad attention but has focused on storage of sensory content. Evidence on short-term maintenance of abstract verbal or categorical information is scarce. This thesis aims to investigate neural representation of WM content at different levels of abstraction. I present here three empirical studies that employed fMRI, multivariate pattern analysis or probabilistic modeling as major methods. The first study identified cortical regions that retained WM content of a script. Native Chinese speakers were asked to memorize well-known Chinese characters which strongly facilitated verbal coding. Results indicated left lateralized language-related brain areas as candidate stores for verbal content. The second and the third studies aimed to test the hypothesis that color is memorized as a combination of the low-level visual representation and the abstract categorical representation. The second study utilized a conventional sensory encoding model and a novel empirical-based categorical encoding model to characterize two sources of neural representations. Color information was decoded in three color-related ROIs: V1, V4, VO1, and notably, an elevation in categorical representation was observed in more anterior cortices. In the third study, the delayed behavioral response was examined, which exhibited a systematic bias pattern; a probabilistic dual-content model was implemented, which produced response patterns highly correlated with experimental results; this confirmed the hypothesis of dual-content mnemonic representations. These studies together suggest a division of labor along the rostral-caudal axis of the brain, based on the abstraction level of memorized contents.

Zusammenfassung

Die Erforschung der neuronaler Grundlagen des Arbeitsgedächtnisses (WM) fand breite Aufmerksamkeit, konzentrierte sich aber auf die Speicherung sensorischer Inhalte. Beweise für die kurzfristige Aufrechterhaltung abstrakter, verbaler oder kategorischer Informationen sind selten. Ziel dieser Arbeit ist die Untersuchung der neuronalen Repräsentation von WM-Inhalten auf verschiedenen Abstraktionsebenen. Ich stelle hier drei empirische Studien vor, in denen fMRT, multivariate Musteranalyse oder probabilistische Modelle als Hauptmethoden eingesetzt wurden. Die erste Studie identifizierte kortikale Regionen, die den WM-Inhalt eines Skripts behielten. Chinesische Muttersprachler wurden gebeten, sich bekannte chinesische Zeichen zu merken, was die verbale Kodierung stark fördern. Die Ergebnisse zeigten links lateralisierte sprachbezogene Hirnareale als Kandidatenspeicher für verbale Inhalte. Die zweite und dritte Studie zielten darauf ab, die Hypothese zu testen, dass Farbe als eine Kombination aus einer visuellen Repräsentation und einer kategorischen Repräsentation gespeichert wird. Die zweite Studie verwendete ein sensorisches Kodierungsmodell und ein empirisch basiertes kategorisches Kodierungsmodell, um jeweils zwei Quellen neuronaler Repräsentationen zu charakterisieren. Farbinformationen wurden in drei farbbezogenen ROIs dekodiert: V1, V4, VO1, und insbesondere wurde eine Erhöhung der kategorischen Repräsentation in vorderen kortikalen Arealen beobachtet. In der dritten Studie wurde die verzögerte Verhaltensreaktion untersucht, die ein systematisches Bias-Muster zeigte; es wurde ein probabilistisches Dual-Content-Modell implementiert, das ein mit den experimentellen Ergebnissen hoch korreliertes Antwortmuster erzeugte; dies bestätigte die Hypothese der mnemonischen Dual-Content Repräsentation. Diese Studien zusammen schlagen eine Arbeitsteilung entlang der rostro-kaudalen Achse des Gehirns, die auf der Abstraktionsebene der gespeicherten Inhalte basiert.

Chapter 1 General Introduction

In this chapter, I start by introducing three types of human memory: sensory memory, short-term memory and long-term memory (section 1.1). This section describes the characteristics of each memory, their differences, and the information transfer between them. In the following section (section 1.2), I introduce the theoretical framework of the architecture of working memory. According to an influential cognitive model from Baddeley and Hitch, working memory consists of four sub-components for control or for storage of information from different modalities. In the last section (section 1.3), I address the ongoing debate over the neural basis of working memory storage and review neuroimaging studies that used multivariate pattern analysis to identify brain regions maintaining memory content.

1.1 The Basics of Human Memory

'The present is object only of perception, and the future, of expectation, but the object of memory is the past.'

(Aristotle, 350 BC; translated by Beare, 2010)

Memory is a crucial capability of the human mind. With memory, we accomplish a diversity of cognitive tasks. For example, we rely on memory to know where to go to have a safe rest, when to plant seeds to produce a good harvest, how to drive a car or do a math calculation, what to avoid eating to stay healthy, and more importantly, to know who we are (typical counterexample: Alzheimer's disease; Carlesimo and Oscar-Berman, 1992). Thus, for centuries, human memory has been an intriguing subject of scientific and philosophical inquiry.

The Multi-Store Model of Human Memory

Human memory is generally considered to be composed of three components (Atkinson and Shiffrin, 1968). One simple way to differentiate between them is to inspect the time limit before decay. Information can be held for the shortest time in *sensory memory* (SM; also called sensory register), longest in *long-term memory* (LTM), and for an intermediate period of time in *short-term or working memory* (STM or WM). The term 'working memory', which allows for manipulation of stored information in addition to short-term memory (Hooker, 1960; Baddeley and Hitch, 1974), is dominantly used in this thesis.

Atkinson and Shiffrin proposed that in the multi-store model (**Figure 1-1**; Atkinson and Shiffrin, 1968), stimulus information firstly automatically reaches SM and resides there very briefly. While most of the information in SM decays and is forgotten, some is transferred to WM through attentional selection. Information in these two stores, however, does not always share the same modality. The model argues that a limited amount of information can be held in WM for a short period of time through rehearsal. Some of the information retained in WM can be transferred to LTM with little conscious awareness (Hebb, 1961; Melton, 1963). LTM can store information for a longer period of time than seconds to minutes, which can further be retrieved and transferred to WM.

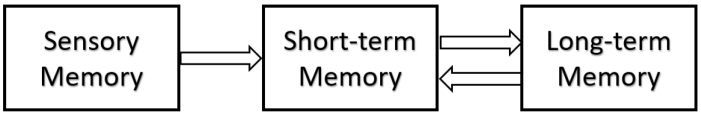

Figure 1-1 The illustration of the multi-store model proposed by Atkinson and Shiffrin in 1968. It describes how information is transferred between three components of human memory (Adapted from Atkinson and Shiffrin, 1968).

This early model is simplistic but captures some essential principles. Later studies exhibit evidence that revise the model in various ways. For example, patient K.F. who suffered brain damage with diminished verbal WM capability was found to have intact visual WM performance (Zlonoga and Gerber, 1986), suggesting WM is composed of distinct sub-stores with separate neural systems. This idea was further extended to suggest /argue that every human memory component can be further divided into multiple sub-stores (Darwin et al., 1972; Baddeley and Hitch, 1974; Graf and Schacter, 1985). How information is transferred between three components has also been investigated It was discovered that the way in which precise information is stored in LTM is not related to how long it is retained but to the processing level

in WM (Craik and Lockhart, 1972). Furthermore, lesion studies show clear evidence against the view that all information reaches LTM through WM, as impairment of WM ability does not necessarily cause damage to LTM performance (Shallice and Warrington, 1970; Baddeley et al., 1988).

Sensory Memory (SM)

Today, SM is generally thought to include multiple sensory sub-stores which maintain information of distinct modalities. For example, visual information is kept in iconic memory store (Sperling, 1960), auditory information is held in echoic memory store (Darwin et al., 1972), and tactile information is maintained in haptic memory store (Bliss et al., 1966). When a sensory stimulus is presented, input information automatically enters SM and can be maintained here for a very brief time up to several hundred milliseconds (Sperling, 1960; Averbach and Coriell, 1961; Estes and Taylor, 1964). SM can maintain information with high precision and high capacity, but is vulnerable to interference (Sperling, 1963; Waugh and Norman, 1965).

Long-Term Memory (LTM)

In contrast to the other two types of human memory, contents stored in LTM can be maintained for a nearly unlimited period of time (Bahrick et al., 1975; Bahrick, 1984). It is also shown to have high precision and high capacity (Standing, 1973). The level of LTM recollection is strongly subject to retrieval cues, and memorized information could appear inaccessible due to different factors such as overwhelmed association network and inappropriate cues (Feigenbaum, 1961; Tulving and Pearlstone, 1966).

LTM is conventionally divided into two categories: *explicit memory* and *implicit memory* (Graf and Schacter, 1985). Explicit memory (also called *declarative memory*) refers to consciously accessible memory that can be expressed explicitly (Graf and Schacter, 1985; Schacter and Graf, 1986). It can be further divided into *episodic memory* which maintains personal experiences and *semantic memory* which holds information about words and concepts (Tulving, 1972, 1989). In contrast, implicit memory (or *procedural memory*) is often implicitly acquired and used for facilitation in motor tasks without conscious recollection (Graf and Schacter, 1985; Schacter and Graf, 1986; Schacter, 1987). For example, the memory of how to move one's legs and arms in coordination in order to ride a bike without consciously thinking about it is implicit memory.

However, where in the brain LTM content is encoded and stored is not completely clarified. It has been found that declarative memory is highly dependent on the hippocampus and surrounding cortices (Scoville and Milner, 1957; Vargha-Khadem et al., 1997), while procedural memory content is encoded and maintained mainly in the basal ganglia (Foerde and Poldrack, 2009).

Evidence shows that sleep facilitates the consolidation of newly acquired LTM information, for both explicit and implicit memory (Jenkins and Dallenbach, 1924; Barrett and Ekstrand, 1972; Plihal and Born, 1997; Stickgold et al., 2000; Maquet, 2001; Fischer et al., 2002; Gais et al., 2002). During sleep, the newly encoded, labile memory of both categories can be quantitatively strengthened (Plihal and Born, 1997; Stickgold et al., 2000; Fischer et al., 2002; Gais et al., 2002; Walker et al., 2003; Ellenbogen et al., 2006; Korman et al., 2007) and qualitatively reorganized to facilitate generalization and to inspire new insights (Wagner et al., 2004; Fischer et al., 2006; Ellenbogen et al., 2007; Diekelmann and Born, 2010). Memory encoding and consolidation processes occur at separate times during wakefulness and sleep to avoid mutual interference which could cause hallucination (Brown and Robertson, 2007; Robertson, 2009; Diekelmann and Born, 2010).

Working Memory (WM)

WM generally refers to the temporary maintenance of information that is no longer present, as well as the manipulation of the memorized content (Baddeley, 2003; Postle, 2006; Zimmer, 2008a). Information can reach WM either through attentional selection from SM or from LTM, and can be held here for a short time span from seconds to minutes (Atkinson and Shiffrin, 1968). It has been widely accepted that internal rehearsal is essential for WM based on findings that articulation rate and memory span are linearly related (Landauer, 1962; Baddeley et al., 1975). However, this view has been contested by recent studies arguing that articulation is not equivalent to the rehearsal process, and other unclarified mechanisms might contribute to WM (Caplan et al., 1992; Cowan et al., 1998; Service, 1998; Hulme et al., 1999; Lovatt et al., 2000; Nairne, 2002).

It is still under debate as to how information is maintained in WM with limited capacity (Luck and Vogel, 2013; Ma et al., 2014). It has been proposed that WM content is retained as a limited number of discrete representations with fixed-resolution (Luck and Vogel, 1997). Classic examples are the magical number seven (Miller, 1956) or four (Cowan, 2001). This ‘slot’ view

suggests an all-or-none memorization of each item, and is supported by conventional paradigms such as the digit span test and the change detection task (Pashler, 1988; Luck and Vogel, 1997; Engle et al., 1999; Rouder et al., 2008). In contrary, others have argued that there is a limited amount of memory resources that can be allocated among multiple representations with variable resolution (Frick, 1988). In contrast to conventional paradigms with discrete numbers of items, the delayed estimation paradigm with continuous feature space was developed to test the 'resource' concept (Wilken and Ma, 2004; Zhang and Luck, 2008). In addition, it has been proposed that a flexible memory resource can be occasionally utilized in slot style, as a combination of two views (Donkin et al., 2016).

WM is crucial for a variety of essential cognitive functions such as reasoning, learning, mathematical calculation, language acquisition and other fluid intelligences (Hitch, 1978; Baddeley, 1986, 2003; Kyllonen and Christal, 1990; Engle et al., 1999). Improvement in WM capacity in childhood is considered a key predictor of development in cognitive abilities (Andrews and Halford, 2002; Jarrold and Bayliss, 2007). WM impairment is commonly seen in neural disorders featuring diminished cognitive functions such as Alzheimer's dementia, ADHD, Parkinson's disease and Huntington's disease (Willcutt et al., 2005; Lee et al., 2010; Liu et al., 2014; Poudel et al., 2015). WM functions also tend to decline in old age (Park et al., 2002; Hertzog et al., 2003). Furthermore, WM can be sub-divided into multiple components (Hooker, 1960; Baddeley and Hitch, 1974), which is discussed in detail in the next section.

To summarize, human memory can be divided into three categories: sensory memory, long-term memory and working memory (Atkinson and Shiffrin, 1968). These three kinds of human memory differ in their memory span and capacity, as well as in the neural mechanisms underlying the encoding and maintenance processes (Ranganath and Blumenfeld, 2005). While working memory is often considered to be linked to temporary electrical activation, long-term memory is formed through lasting neuronal change (Hebb, 1949). Next, while the encoding process of working memory depends strongly on attention (Atkinson and Shiffrin, 1968), information encoding in sensory memory is automatic and exhibits only weak dependence on attention (Sperling, 1960; Persuh et al., 2012). The neural mechanism of memory is complex and the neuroscientific understanding of how memory is stored is far from settled. The main research objective of this thesis is to investigate the neural mechanism of working memory, t focusing on identifying the cortical storage sites of working memory contents at different levels of abstraction.

1.2 Baddeley's WM Model

To understand the composition of working memory, various cognitive models have been proposed (Cowan, 2001, 2012; Oberauer et al., 2012; Oberauer and Lin, 2016). Among them the most influential one was proposed by Baddeley and Hitch in 1974, suggesting a division of working memory into three main components: the central executive, the visuospatial sketchpad and the phonological loop (Baddeley and Hitch, 1974). Later, this model was further extended by adding a fourth component: the episodic buffer (Baddeley, 2000).

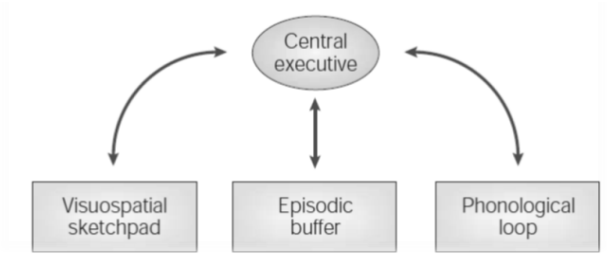

Figure 1-2 Illustration of Baddeley and Hitch's model of working memory, which is composed of four main components (adapted from Baddeley, 2003).

Central Executive

The model proposes that the *central executive* serves as a control and regulation center and is responsible for directing attention to relevant information and for coordinating information transfer as well as controlling other cognitive processes (Baddeley and Hitch, 1974; Wongupparaj et al., 2015). Despite its essential role, the mechanisms of the central executive are little understood.

Visuo-Spatial Sketchpad

The visuo-spatial sketchpad is dedicated to the maintenance as well as manipulation of visual and spatial information (Baddeley and Hitch, 1974). Different object features from the same dimension might compete for a limited amount of memory slots or resources (Luck and Vogel, 1997; Vogel et al., 2001). The visuo-spatial sketchpad can be subdivided, in analogy to a

phonological loop, into a *visual cache* for storage and an *inner scribe* for dynamic retrieval and rehearsal processes (Logie and Logie, 1995).

Episodic Buffer

The episodic buffer can integrate information from subsidiary working memory components as well as from long-term memory in time sequence (Baddeley, 2000). It bridges the link not only between working memory systems, but also between working memory and other human memory (Baddeley, 2011). The central executive can affect the formation of episodic memory, and conscious awareness is key to retrieving information from episodic buffer (Baddeley, 2011). Further, through episodic buffer, one may consciously access information held in the phonological loop or the visuo-spatial sketchpad (Baddeley et al., 2011).

Phonological Loop

The phonological loop consists of two sub-units and serves retention of auditory or verbal information (Baddeley and Hitch, 1974). The *phonological store* retains memory traces for a short period of time before rapid decay, and the *articulatory rehearsal process* prevents memory storage from decaying (Baddeley and Hitch, 1974). The two components work in cooperation, while the former acts as an 'inner ear' and stores encoded information, the latter acts as an 'inner voice' and revives the memory trace by repeatedly retrieving and articulating it (Vallar and Papagno, 2002; Baddeley, 2003). In addition, visually presented stimuli can be visually analyzed and temporarily retained in the visual memory store before being converted into phonological information (grapheme-to-phoneme conversion) and transferred to phonological loop through articulatory rehearsal (Vallar and Papagno, 2002; Baddeley, 2003) .

This phonological system has a limited capacity, which can be explained by the nature of the articulation process in real time (Baddeley, 2003). For example, to memorize a large number of items, such a long period of time is needed for one round, that the first item might be forgotten by the time the last item is articulated. The following factors may influence WM capacity of verbal and acoustic information: (1) *Word-length effect* (Baddeley et al., 1975, 1984). The WM capacity of words is inversely related with the word length. For example, the phonological loop exhibits a smaller memory span for long words with many syllables than short ones with few syllables. (2) *Phonological similarity effect* (Conrad, 1964; Conrad and Hull, 1964; Wickelgren, 1965; Baddeley, 1966). A set of stimuli with similar pronunciations are difficult to memorize

compared to words with dissimilar sound. Heavy phonological similarity can cause abandoning the phonological loop and switching to visual or other type of coding. (3) *Articulatory suppression effect* (Murray, 1968; Levy, 1971; Peterson and Johnson, 1971; Baddeley et al., 1975, 1984; Vallar and Papagno, 2002). Continuously voicing irrelevant contents can lead to reduced verbal WM capacity, as well as eliminate the word-length effect and phonological similarity effect.

Baddeley and Hitch's model, which proposes multiple buffers in working memory, has received enormous attention in the field of psychology. But to ascertain how working memory functions, additional neuroscientific evidence is needed.

1.3 The Debate over the Neural Basis of WM Storage

An important question to ask in neuroscience is: what is the basic neural substrate for the maintenance of different types of information in working memory? Since the 1970s, a number of studies have been conducted using single unit electrophysiological approaches to examine the monkey brain, and it was found that individual neurons in prefrontal cortex (PFC) exhibited sustained neural activity over the delay period (Fuster and Alexander, 1971; Fuster, 1973; Niki and Watanabe, 1976; Watanabe, 1981; Fuster et al., 1982; Quintana et al., 1988; Funahashi et al., 1989, 1990). Later in 1995, Goldman-Rakic proposed to integrate these neuroscientific findings in PFC with the psychological model from Baddeley and Hitch (Baddeley and Hitch, 1974; Baddeley, 2000), resulting in a widely influential standard model of working memory. This standard model states that specialized systems centralized in the lateral prefrontal cortex (lPFC) are responsible for both the storage and manipulation of working memory contents (Goldman-Rakic, 1995). Lesion studies provided supporting evidence that impairment in lPFC diminished working memory performance (Jacobsen, 1935; Gross, 1963; Anon, 1964; Goldman and Rosvold, 1970; Petrides and Milner, 1982; Funahashi et al., 1993; Ptito et al., 1995).

However, this 'centralized' perspective has been challenged by a large number of recent empirical studies, and as it often happens in science, the widely influential standard model might need a revision (Postle, 2006).

Since the development of multi-voxel pattern analysis (MVPA) for feature-specific decoding in 21st century (Haxby et al., 2001; Haynes and Rees, 2005a; Kamitani and Tong, 2005a; Norman et al., 2006; see section 2.3.2), multiple neuroimaging studies have been conducted to examine working memory storage in the human brain. Contrary to what the standard model claims, distributed brain areas were found to retain stimulus-specific representations of various types of stimuli over the delay period. Studies on visual features identified orientation information retained in the early visual cortex, the posterior parietal cortex, frontal eye fields and the lateral prefrontal cortex (Harrison and Tong, 2009; Serences et al., 2009, 2009; Sneve et al., 2012; Albers et al., 2013; Ester et al., 2013, 2015; Pratte and Tong, 2014; Bettencourt and Xu, 2016); color information maintained in V1 (Serences et al., 2009); and complex shapes retained in the lateral occipital complex, the posterior parietal cortex, and frontal eye fields (Christophel and Haynes, 2014a). Memory storage of auditory information was localized in the auditory cortex (Linke and Cusack, 2015; Kumar et al., 2016). Further, motion flow patterns were decoded from hMT+ (Emrich et al., 2013; Christophel and Haynes, 2014a), spatial locations were decoded from frontal eye fields (Jerde et al., 2012), and complex visual patterns were decoded from the posterior parietal cortex (Christophel et al., 2012; Christophel and Haynes, 2014a). An alternative idea was thus proposed, arguing for a coordinated recruitment of distributed regions covering the neocortex for working memory maintenance (Fuster, 1995; Postle, 2006; Zimmer, 2008a; Christophel et al., 2017).

It should be noted, that some recent works found feature-specific working memory content of object (Lee et al., 2013), orientation (Ester et al., 2015) and auditory information (Kumar et al., 2016) stored in both the lateral prefrontal cortex and posterior sensory regions. These recent studies raised the question of whether the dual representations of the memorized information are redundant. An interesting explanation refers to a labor division between anterior cortices and posterior sensory regions based on the abstraction level of the memorized content (Christophel et al., 2017). This perspective extends the alternative ‘distributed’ view and is the key hypothesis of this thesis (further discussed in 0).

Although research on the neural basis of working memory has gained widespread attention in the last decades, empirical work has mainly focused on sensory forms of working memory. Evidence on short-term maintenance of abstract verbal or categorical information is scarce.

Relevant work using verbal stimuli either examined the maintenance of two roman letters but could not ascertain the modality of the memorized content (Polanía et al., 2011), or investigated

the contrast between language and non-verbal contents (Lewis-Peacock et al., 2012; Yue et al., 2018), and thus lacking specificity for verbal WM contents. The first study (Chapter 3) in this thesis aims at 1) directly investigating working memory storage of verbal material and thus adding the missing evidence to the field; 2) providing evidence to address the ongoing debate on the cortical localization of working memory storage; 3) delivering neuroscientific evidence to address Baddeley's model of the phonological loop. This study utilized fMRI and searchlight-based multivariate pattern analysis (Chapter 2) to identify cortical regions that hold feature-specific language content over the delay period.

The second and third studies (Chapter 4 and 0) test the hypothesis that color working memory is realized through a combination of the low-level visual representation and the abstract categorical representation. In Chapter 4, the dual-content mnemonic representation in sensory regions is assessed by two types of color-selective encoding models together with multivariate pattern analysis. Low-level visual and high-level categorical neural representations of color working memory are respectively characterized by a conventional cosine-shaped encoding model and an empirical-based categorical encoding model. In addition, the color vision is tested and contrasted with the color working memory. In 0, the hypothesis is tested by examining response patterns in behavioral data as well as by implementing a dual-content probabilistic model.

Chapter 2 Analysis of fMRI Data

This chapter aims at providing relevant knowledge about the major methods used in this thesis. I start with introducing basic information about functional magnetic resonance imaging (fMRI; section 2.1) and preprocessing procedures (section 2.2). Then, a long section is dedicated to introducing different fMRI analysis methods (section 2.3). Univariate analysis (section 2.3.1) is introduced briefly, while multivariate pattern analysis (MVPA) is described in four subsections (section 2.3.2). Finally, multivariate and univariate analyses are compared, and their differences are discussed (section 2.3.3).

Although it is not the main purpose of this thesis to extensively understand fMRI physics or fMRI analysis methods, it is crucial to be aware of the capacities and limitations of the techniques, in order to interpret the findings from fMRI studies. A critical part of the fMRI analysis in this thesis is to decode stimulus-selective information based on brain activity patterns via MVPA. In section 2.3.2, I first introduce the general background and concepts of MVPA (section 2.3.2.1), followed by searchlight-based brain mapping and region of interest analysis (section 2.3.2.2); next, I describe in details the computation of a MVPA method used in this thesis: cross-validated MANOVA (section 2.3.2.3); and finally the inverted encoding model, which utilizes encoding basis functions to characterize selective neural responses to stimuli, is introduced (section 2.3.2.4).

2.1 Basics about fMRI

Magnetic resonance imaging (MRI) is a widely used non-invasive method for examining the anatomy and physiological processes of the human body. This approach relies on a number of sophisticated physical mechanisms, especially the nuclei magnetic resonance (NMR) properties of protons in magnetic fields (Schild, 1990; Huettel et al., 2014). In order to form the anatomical image of the body, it measures the *radio-frequency (RF)* signals sent from protons in hydrogen atoms that compose the water and fat molecules in the body.

The *functional magnet resonance imaging (fMRI)* is influenced by the oxygen level in the blood, from which brain activity can be inferred. This inference is based on the assumption that the cerebral blood increases its flow and volume to deliver nutrients like oxygen to surrounding active brain tissues, which consumes energy. More specifically, oxygen is bound to the hemoglobin molecule for transport in the blood, and a decrease in oxygen level due to neural activity leads to a ratio change between *oxygenated hemoglobin (Hb)* and *deoxygenated hemoglobin (dHb)*. While Hb is diamagnetic, dHb is paramagnetic and changes magnetic susceptibility between blood vessel and neighboring brain tissues. This susceptibility difference induces change in the proton resonance frequency of hydrogen atoms in water molecules, and thus generates the *blood oxygenation level dependent (BOLD)* contrast (Ogawa et al., 1990). The increase in paramagnetic dHb often leads to a decrease in BOLD signal.

Although the BOLD signal can reflect neural activity (Logothetis et al., 2001), the coupling between them is not simple and direct (secondary consequence). A complex temporal change is typically observed in the BOLD signal in response to an external stimulus (Glover, 1999). When the stimulus is presented for a very short time, the hemodynamic BOLD response typically looks like **Figure 2-1**. The neural activity in a brain area responding to a stimulus can lead to increased oxygen consumption and thus increased dHb quantity, resulting in an initial dip of the BOLD signal. This is compensated for by increasing cerebral blood flow and volume, which causes a strong increase in the Hb supply, and thus a positive BOLD response. After reaching the peak (typically in 5 s), the BOLD signal drops below the baseline, which is called the post-undershoot, and then returns slowly back to the resting state (Buxton et al., 1998).

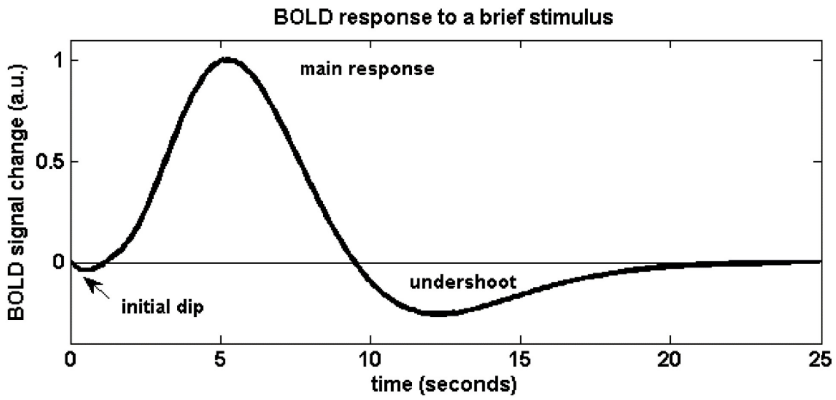

Figure 2-1 Illustration of the typical BOLD change in response to a stimulus starting at time 0 (adapted from Barth and Poser, 2011). After an initial dip the BOLD response rises and reaches the peak in approximately 5 s. The return back to the baseline is typically preceded by an undershoot of the BOLD signal.

Since its first proposal and usage in 1992 (Bandettini et al., 1992; Kwong et al., 1992; Ogawa et al., 1992), fMRI has been widely employed to examine human neural activity in vivo in a non-invasive way. Compared to other non-invasive functional neuroimaging approaches such as EEG and MEG, fMRI exhibits a high spatial resolution but a relatively low temporal resolution.

2.2 fMRI Preprocessing

A series of preprocessing procedures are conventionally conducted to correct artifacts in the fMRI data or to facilitate further analysis. This section provides a short overview of preprocessing procedures conducted in the thesis (Poldrack et al., 2011) using SPM (SPM8 in Chapter 3, SPM12 in Chapter 4; Friston K. J. et al., 1994). In practice, some operations (e.g. normalization and smoothing for group analysis) are performed after MVPA to maintain the fine spatiotemporal properties of the BOLD signal, and thus to maximize the sensitivity of the MVPA analysis.

Realignment

Firstly, fMRI images are spatially realigned to a reference image (often the first fMRI image of the subject) to correct for the head motion. A least squares approach is applied to perform a six-parameter rigid-body transformation (with no change in head size or shape) that includes translational and rotational corrections along each coordinate of the three-dimensional space (Friston et al., 1995). However, it is not always possible to eliminate movement artifacts through realignment; for example, abrupt head movement in the middle of a scan can cause hard-to-correct changes in the image intensity (Poldrack et al., 2011). In a worst-case scenario, the data of a whole experimental run or subject need to be discarded.

Coregistration

The structural image with a relatively high spatial resolution is coregistered to the functional EPI image (e.g. the first EPI scan). This procedure is performed to compensate for the relatively low spatial resolution of the functional image, which can make it difficult to discriminate its anatomical boundary.

Unified Segmentation

The unified Segment function merges three procedures: spatial normalization, tissue partition and bias field correction (Ashburner and Friston, 2005). A segmented brain image is spatially warped to match a template brain, classified into separate tissue compartments, and corrected for variation in intensity. In our studies, the coregistered structural image is segmented in preparation for normalization.

Normalization

Brain data from different subjects are statistically tested as a group to draw inference from the population. However, the anatomical variation in brain shape and size makes the group-level analysis difficult. To solve this problem, brain images of individual subjects are spatially warped into a standard space (e.g. in MNI space), which is called normalization. The coregistered and segmented structural image is employed as the deformation field for the procedure. In this thesis, normalization is applied after MVPA to maintain the fine spatiotemporal properties, and thus to maximize the analysis sensitivity.

Smoothing

Spatial smoothing is often conducted to reduce noise signal, to increase statistical power, as well as to account for individual differences (Friston et al., 2000). Smoothing is performed after MVPA by convolving with a Gaussian kernel. The optimal full-width at half maximum (FWHM) of the Gaussian kernel can vary depending on multiple factors (Mikl et al., 2008). Based on previous lab experience with working memory data (Christophel et al., 2012; Christophel and Haynes, 2014a), the Gaussian kernel with a FWHM of 5 mm in all directions (x, y and z) is used in this thesis.

2.3 fMRI Analysis

2.3.1. Univariate Analysis

To estimate brain activity in response to experimental conditions, a *general linear model (GLM),* which relates observed BOLD signals to relevant experimental variables (Friston et al. 1994),

is applied. In cases where multiple voxels are considered, the model can be written as an equation of matrices (also called multivariate general linear model MGLM):

$$Y = X\beta + \varepsilon, \qquad (2\text{-}1)$$

Where Y stands for measured hemodynamic signals as a two-dimensional matrix (rows: scans; columns: voxels), X depicts the design matrix (rows: scans; columns: regressors), ß represents the to-be-estimated parameter matrix (rows: regressors; columns: voxels), and ε depicts the error matrix, which is assumed to be independent and normally distributed (rows: scans; columns: voxels). In other words, the observed hemodynamic response can be modeled as the linear sum of multiple weighted regressors plus the error in each of the multiple examined voxels. In order to estimate the optimal parameter matrix ß with a minimal residual ε, linear regression (like least-squares or Bayesian approaches) is often conducted.

The design matrix X relates BOLD signals to relevant experimental events, and thus frames the hypothesis to be tested. To characterize the shape of the BOLD impulse response to an experimental event, a canonical *hemodynamic response function (HRF)* is typically utilized. HRF characterizes the ideal hemodynamic response to a delta pulse stimulus with no noise. Alternatively, a *finite impulse response (FIR)* set can be utilized to characterize the BOLD impulse response of any shape. The FIR set consists of a set of basis functions, and models a series of successive time units (Henson et al., 2001). The BOLD response to simple tasks with short durations (e.g. pressing a button or perceiving a stimulus) can often be modeled by a canonical HRF, whereas the brain response to complex tasks with prolonged durations can be better captured by a FIR set (e.g. memorizing a stimulus for a short period of time). By convolving stimulus onset vectors (stick function) with a HRF or FIRs, regressors for an event-related experiment can be acquired (Henson and Friston, 2016).

2.3.2. Multivariate Pattern Analysis

2.3.2.1. MVPA in General

Since the last decade, a new type of fMRI analysis approach has been increasingly utilized that evaluates content-specific information by examining activity patterns of an ensemble of voxels. This type of analysis on multiple voxels is called *multi-voxel/multivariate pattern analysis*, and

MVPA in short (Norman et al., 2006; Haxby, 2012). In some occasions it is also referred to in other terms such as information-based imaging (Kriegeskorte et al., 2006) or decoding (Haynes and Rees, 2006).

The development of the MVPA method likely started with a fMRI study that predicted face and object categories by correlating neural response patterns of multiple voxels to experimental conditions (Haxby et al., 2001). This was followed by another study (Cox and Savoy, 2003) that employed linear discriminant analysis (LDA) and support vector machine (SVMs; Cortes and Vapnik, 1995) to reliably classify object categories with above-chance accuracy. In the next few years following that study, multiple fMRI studies utilized activity patterns of a set of voxels to reliably predict visually perceived content such as orientation information (Kamitani and Tong, 2005a), as well as current cognitive states (Mitchell et al., 2003), consciously invisible stimuli (Haynes and Rees, 2005a) and subjective conscious experience (Haynes and Rees, 2005b). Notably, it was found that MVPA could reveal feature-specific representation in a brain area where no effect is detected with the univariate analysis (Haynes and Rees, 2005a; Kamitani and Tong, 2005a).

Conventionally, MVPA quantifies the feature-specific information based on how well a classifier could classify between different experimental conditions (Haynes and Rees, 2006). A *SVM* is often employed to classify multivariate data (Cortes and Vapnik, 1995; Cox and Savoy, 2003). The classifier is first trained on part of the experimental data to distinguish conditions based on multi-voxel activity patterns. Then the trained classifier is tested on the remaining novel data for condition classification. A *n-fold cross-validation* (n-fold CV; Duda et al., 2000) approach is often used, where the data is divided into n equal-sized parts, and the procedure is repeated n times until every part is used once for validation. This approach utilizes all acquired brain data for both training and testing while avoiding overfitting or selection bias issues (Cawley and Talbot, 2010). The results from n repetitions are averaged to estimate the final classification accuracy.

An alternative MVPA approach that quantifies the amount of multivariate covariance specified by a contrast matrix in units of the multivariate error covariance has been developed (Allefeld and Haynes, 2014). This approach, *cross-validated MANOVA (cvMANOVA)*, is based on multivariate analysis of variance (MANOVA; see Timm, 2002), but utilizes a leave-one-session-out cross-validation approach to avoid biases, which is equivalent to the n-fold cross-validation when n equals the session number. Its result, *pattern distinctness D,* is an unbiased

estimate of the explained multivariate variance. When only two experimental conditions are considered, D is equivalent to the Mahalanobis distance (Mahalanobis, 1936). In short, cvMANOVA can be understood as estimating the generalized squared distance between conditions based on multivariate data. It is argued to be a more direct method for examining feature-selective content, because it directly measures the degree to which activity patterns of experimental conditions differ (Hebart and Baker, 2017). In contrast, classifier-based MVPA relies on multiple factors such as the selection of the classifier, the algorithm as well as relevant parameters, and is a rather indirect approach (Allefeld and Haynes, 2014). For these reasons, cvMANOVA is employed as the major MVPA approach in this thesis.

2.3.2.2. ROI-based and Searchlight-based Analysis

Analysis of fMRI data can be performed within the pre-identified *region-of-interest (ROI).* There are different ways to define ROIs, and ROI can be based on both anatomical and functional criteria. Because region boundaries can vary across subjects, individual ROIs are often defined. A *retinotopic mapping* approach that utilizes the traveling-wave approach with ring and wedge stimuli is often employed (Sereno et al., 1995; DeYoe et al., 1996; Engel et al., 1997b; Wandell et al., 2007) to identify visual field maps in visual areas (such as lateral occipital cortex including LO1, LO2, hMT+; ventral occipital cortex including V4, VO1, VO2; and dorsal visual cortex including V3A, V3B). However, retionotopic mapping has several practical limitations; for example, it depends largely on utilized stimuli and parameters, and it demands multiple fMRI sessions for each subject, thus making it costly (Wandell et al., 2007). These limitations were addressed by a study which estimated probabilistic maps in visual areas based on empirically acquired topographic maps of 53 subjects (Wang et al., 2015). These probabilistic maps of visual topographic areas are utilized to study color working memory in this thesis (Chapter 4).

The introduction of the *searchlight* method to content-specific analysis allows multivariate analysis on the whole brain level (Kriegeskorte et al., 2006). It liberates us from the restrictions of early MVPA studies that were from within ROIs in the visual cortex (Haxby et al., 2001; Haynes and Rees, 2005a, 2005b; Kamitani and Tong, 2005a, 2005b) and temporal cortex (O'Toole et al., 2005). The basic idea is to combine all voxels within a small region for multivariate analysis, and to search through the whole scanning volume for content-specific

information. For continuous brain mapping, one can analyze every voxel by examining activity patterns of the surrounding spherical ensemble of voxels, and repeat this for all voxels throughout the brain (Kriegeskorte et al., 2006). We employed the spherical volume with a radius of five voxels in this thesis (**Figure 2-2a)**. This searchlight-based MVPA approach enables us to examine the entire brain for feature-specific patterns without any pre-assumption about the effect location and is thus widely used to analyze fMRI studies.

2.3.2.3. Computation of cvMANOVA

The cross-validated MANOVA (cvMANOVA) approach estimates the amount of multivariate variance specified by the contrast matrix in relation to the error variance, in order to assess whether multi-voxel activity patterns carry information that can differentiate between variations of the experimental feature (Allefeld and Haynes, 2014). It can be combined with the searchlight approach for the whole brain mapping (Kriegeskorte et al., 2006), or with the ROI analysis for investigation within specific brain regions (Wang et al., 2015). In this section, the computation of the searchlight-based cvMANOVA is described with more detail in three steps (Allefeld and Haynes, 2014).

Firstly, a searchlight approach is employed for brain mapping (**Figure 2-2a**). To assess the information carried by a single voxel V_i, a set of voxels within the spherical volume surrounding this voxel V_i are extracted and examined jointly. This process is repeated voxel by voxel throughout the brain, with a fixed spherical radius of five voxels in this thesis. Secondly, a multivariate general linear model (MGLM) is constructed to estimate the regressor-specific multivariate patterns (**Figure 2-2b**). The measured BOLD signal Y (rows: fMRI scans; columns: voxels) is modeled by a standard design matrix X (rows: scans; columns: regressors). The optimal parameter matrix ß (rows: regressors; columns: voxels) is estimated by minimizing model errors ε (rows: scans; columns: voxels). As an example, here we consider only two voxels within the sphere and three experimental conditions (**Figure 2-2c**). Thirdly, the pattern distinctness D is computed as unbiased estimates of the multivariate covariance (between-class covariance) specified by a contrast matrix (**Figure 2-3**) in relation to the error covariance (within-class covariance; **Figure 2-2c**).

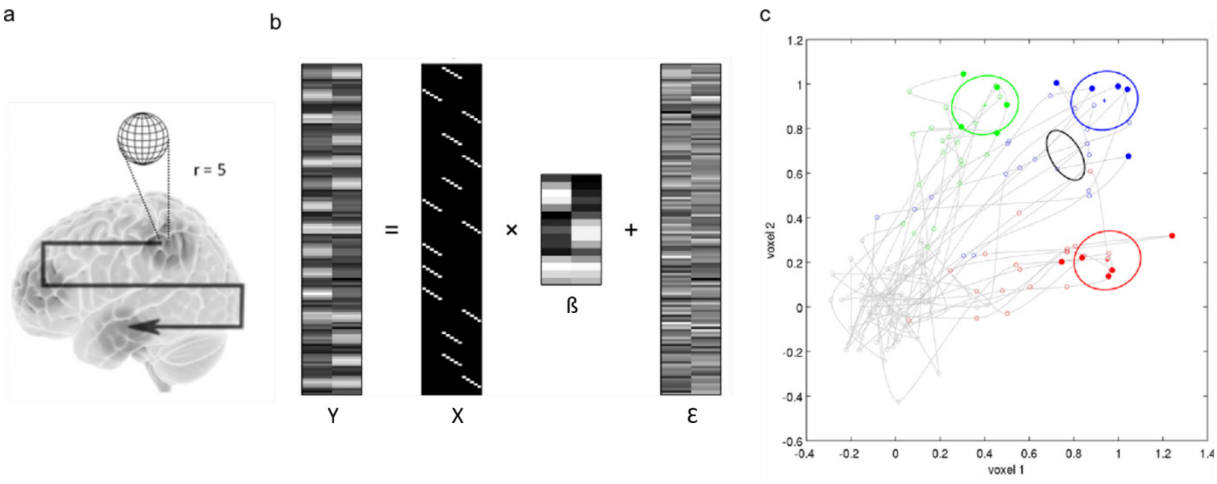

Figure 2-2 Illustration of the searchlight-based cvMANOVA approach (figure partly provided by Dr. Carsten Allefeld). **a)** Searchlight-based brain mapping. For a given voxel, the BOLD signal of all voxels within the spherical volume surrounding this voxel are extracted and analyzed jointly. This process is repeated voxel by voxel throughout the brain. Two voxels within the searchlight sphere, which has a radius of five voxels, are illustrated. **b)** A multivariate general linear model is fitted using a design matrix X (rows: scans; columns: regressors) to estimate the regressor-specific multivariate patterns. The data matrix Y (rows: fMRI scans; columns: voxels) denotes the BOLD signal measured over a 4.5 min working memory session at a TR of 2 s (135 scans) in two voxels. The 10 s delay in each trial is modeled by 5 FIR regressors, resulting in the estimation of 15 regressors for three experimental conditions. The parameter matrix ß (rows: regressors; columns: voxels) is estimated by minimizing the error matrix ε (rows: scans; columns: voxels). **c)** The pattern distinctness D can be computed as the unbiased estimates of the multivariate covariance (between-class covariance) specified by a contrast matrix in comparison to the error covariance (within-class covariance). The markers denote the measurements of the data matrix Y; the gray line represents the BOLD signal trajectory in two voxels; the filled markets in red, blue and green respectively highlight the 3rd of five FIR regressors in each of three experimental conditions; the black ellipse denotes the between-class covariance and the colored ellipses indicate the within-class covariance.

More details about the third step (**Figure 2-2c**) are explained in the following paragraphs, based on the work by Allefeld and Haynes (Allefeld and Haynes, 2014). For a working memory task with three experimental conditions and a 10 s delay, a contrast matrix C is defined to specify contrasts between regressors (**Figure 2-3**). The 10 s delay in each trial is modeled by five 2 s long FIR regressors, and three conditions lead to an estimation of altogether 15 regressors for every voxel. Each pair of neighboring conditions are contrasted in (1, -1) scale for a given FIR bin, resulting in altogether 10 contrasts for 5 FIR bins.

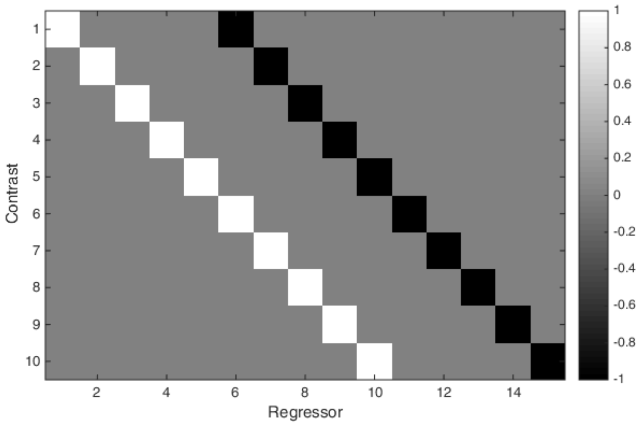

Figure 2-3 The transposed contrast matrix C' between three experimental conditions in a typical working memory task with a 10 s delay.

The contrast matrix C is used to specify an effect of interest that allows the separation of the parameter matrix ß into two parts: a part corresponding to the effect $ß_\Delta$ and a part for the null hypothesis $ß_0$:

$$ß = ß_\Delta + ß_0, \qquad (2\text{-}2)$$

with $ß_\Delta = CC^- ß$ and $ß_0 = ß - ß_\Delta$ ($^-$ refers to the pseudo-inverse calculation). Similar to univariate *analysis of variance (ANOVA)* testing the variance explained by an effect in comparison to the error variance, cvMANOVA compares the between-class covariance with the within-class covariance. While the *between-class covariance* (black ellipse in **Figure 2-2c**) refers to the multivariate covariance explained by the difference between experimental conditions as instructed by the contrast matrix C (**Figure 2-3**):

$$\frac{1}{n} ß_\Delta' X' X ß_\Delta, \qquad (2\text{-}3)$$

The *within-class covariance* (colored ellipses in **Figure 2-2c**) denotes the error covariance around the regressor-specific patterns within each experimental condition:

$$\frac{1}{n} \langle \varepsilon' \varepsilon \rangle, \qquad (2\text{-}4)$$

(where $'$ denotes the matrix transpose and $\langle\ \rangle$ denotes the expectation value). Because between-class and within-class multivariate covariance matrices are of the same size (rows: voxels;

columns: voxels), the comparison between them can be represented by a scalar variable, the *pattern distinctness* D:

$$D = \mathrm{trace}(ß'_{\Delta} X' X ß_{\Delta} \langle \varepsilon' \varepsilon \rangle^{-1}). \quad (2\text{-}5)$$

D estimates the amount of multivariate covariance (Cohen, 1982) explained by the effect, calculated relative to the multivariate error covariance. In the case of a contrast between two classes of experimental events, D equals to the Mahalanobis distance Δ (Mahalanobis, 1936):

$$D = \frac{1}{4}\Delta^2. \quad (2\text{-}6)$$

Therefore, pattern distinctness D can also be seen as a generalized squared distance.

In order to avoid overestimating accuracy of the classifier, the *leave-one-session-out cross-validation* is used to acquire an unbiased estimate of the pattern distinctness. The data consisting of n runs are split into n folds ($k = 1,2, \dots n$). While one run is used for testing (lth run), all the remaining runs are used to train the model. This process is repeated until each run is used exactly once for testing. The pattern distinctness in the lth fold is estimated by

$$\widehat{D}_l = \mathrm{trace}(H_l E_l^{-1}), \quad (2\text{-}7)$$

with

$$H_l = \sum_{k \neq l} \left\{ \widehat{ß}'_{\Delta} \right\}_k \left\{ X' X \widehat{ß}_{\Delta} \right\}_l \quad (2\text{-}8)$$

$$\text{and} \quad E_l = \sum_{k \neq l} \{ \hat{\varepsilon}' \hat{\varepsilon} \}_k. \quad (2\text{-}9)$$

Then the average value across n folds can be computed, thus providing an almost unbiased estimate, which can be further corrected to obtain the unbiased estimate of the explained multivariate variance D. If the true amount of explained variance is 0, estimated values of D are distributed around zero, meaning negative values in approximately half cases. A statistical parametric map of pattern distinctness - SPM$\{D\}$ is obtained by implementing searchlight-based brain mapping. This map reflects the amount of effect-specific information contained in the multi-voxel patterns throughout the brain.

2.3.2.4. Inverted Encoding Model

To examine feature-specific representation of stimulus with continuous feature space (e.g. color, orientation, spatial position), to which neurons respond selectively, an *inverted encoding model (IEM)* can be utilized. It consists of two stages: the encoding stage, which addresses each single voxel individually, and the inverted encoding stage, which takes the activation pattern of multiple voxels into account, making IEM a multivariate analysis (Sprague et al., 2014).

Feature Selectivity and Encoding Basis Functions

The basic assumption of IEM is that neurons respond selectively to a specific feature (e.g. color, orientation, spatial position) with different preferences (Brouwer & Heeger, 2009). Furthermore, it is assumed that a linear relationship exists between the overall voxel response and the sum response of all neurons in that voxel (Brouwer & Heeger, 2009).

To characterize the feature selectivity of neurons, encoding *basis functions (also called channels or filters)* with selective responses to the experimental feature are utilized (Brouwer & Heeger, 2009). The basis functions allow a one-to-one and invertible transformation from the feature space to the channel space. A set of basis functions that respond to different variations of the experimental feature is employed. The selective response of a neuron as well as of a voxel can be modeled as the weighted sum of all channels. Multiple previous studies employed a half-wave rectified cosine function with a high power as the basis function in IEM analysis (Sprague and Serences, 2013; Ester et al., 2015). The cosine function is half-wave rectified to avoid a negative response and to simulate the rare spontaneous firing of cortical neurons. This can also be achieved by raising the cosine function to a power of an even number. Furthermore, the function is raised to a high power to make the channel narrower and more selective. The shape of this cosine-shaped channel is similar to a normal distribution and approximates the shape of the single-unit tuning function of cortical neurons (Brouwer and Heeger, 2009; Ester et al., 2013).

More specifically, two types of cosine-shaped basis functions have been utilized to characterize different features. In order to model selective neural response to spatial positions, a set of spatial channels spanning the spatial space are used (Sprague and Serences, 2013; Sprague et al., 2014). A typical spatial channel can be written as a function of distance:

$$f(r) = 0.5 + 0.5\cos\left(\frac{r\pi}{s}\right)^n \; for\; r < s, or\; 0 \text{ elsewhere.} \qquad (2\text{-}10)$$

Where r refers to the distance from the channel center, s refers to a size constant depicting the shortest distance from the channel center to positons where the function equals 0, and n is the raised power. Often, a set of 36 spatial basis functions raised to the power of 7 is used to model the voxel response (**Figure 2-5a**; Sprague and Serences, 2013; Sprague et al., 2014).

In contrast, in order to characterize selective neural response to orientation, a set of 'steerable filters' is utilized (Freeman and Adelson, 1991; Ester et al., 2013, 2015). The linear combination of these filters can represent arbitrary orientation value. A typical 'steerable filter' is written as:

$$f(x) = 0.5 + 0.5\cos(x - \mu)^n. \qquad (2\text{-}11)$$

Where μ stands for the center of the basis function, x stands for a feature value, and n stands for the raised power. This type of channel can be applied to further represent the circular space of color. In practice, the power value n is often chosen based on the number of basis functions (Ester et al., 2015; https://github.com/tommysprague/IEM-tutorial). For example, with 6 basis functions the cosine function is raised to the power of 6 (**Figure 2-4b**).

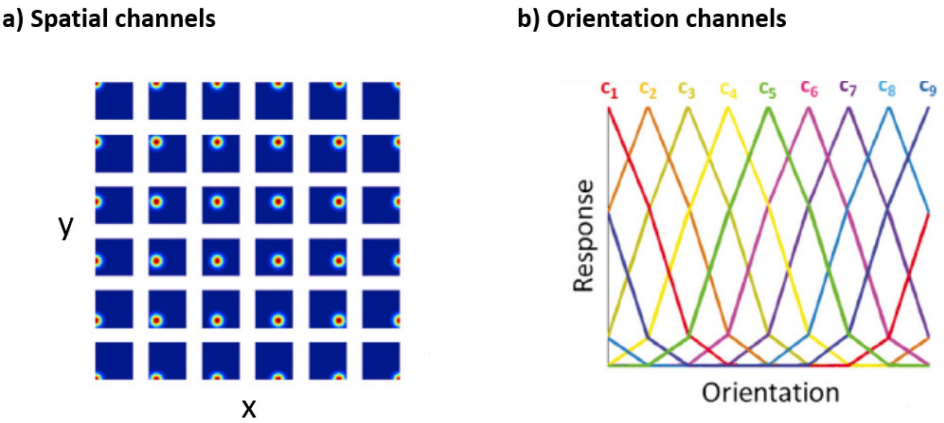

Figure 2-4 Illustration of spatial and orientation basis functions. **a)** A set of 36 spatial channels is utilized to model the neural selectivity in response to spatial information (figure adapted from Sprague and Serences, 2013). **b)** A set of 9 orientation channels is used to simulate the selective neural response to orientation information (figure adapted from Ester et al., 2015).

Computation of IEM

The IEM analysis consists of two stages: encoding and inverted encoding. Accordingly, the data is divided into two subsets (Brouwer & Heeger, 2009): one subset is used to train the model and to estimate the weights (training subset: B_1 and C_1), while the other subset is utilized to test the model and to reconstruct the stimulus feature (testing subset: B_2 and C_2).

In the training stage, the voxel response is modeled by a set of basis functions:

$$B_1 = WC_1, \qquad (2\text{-}12)$$

where B_1 $(m \times n)$ refers to the part of BOLD signal used for training, C_1 $(k \times n)$ stands for channel responses of the training session, and W $(m \times k)$ is the weight matrix (Brouwer and Heeger, 2009; Sprague et al., 2014). Here m represents the number of voxels, n represents the number of trials, and k represents the number of channels. The goal of the training phase is to estimate the weight between the channel space and the brain activity space. In every individual voxel the channel weight is estimated using a general linear model (GLM) and the least square regression approach:

$$\widehat{W} = B_1 C_1^T (C_1 C_1^T)^{-1}, \qquad (2\text{-}13)$$

where $\widehat{W}$ $(m \times k)$ stands for the estimated weight matrix based on the training subset of brain data and channel responses (Brouwer and Heeger, 2009).

In the second stage, the estimated weight of multiple voxels is used to predict the channel response in the testing session. In other words, the mapping is performed across voxels from the observed BOLD signal to the channel response (Sprague et al., 2014). The channel response $\hat{C}_2$ $(k \times n)$ is estimated by:

$$\hat{C}_2 = (\widehat{W}^T \widehat{W})^{-1} \widehat{W}^T B_2, \qquad (2\text{-}14)$$

where B_2 $(m \times n)$ refers to the part of the BOLD signal used for testing (Brouwer and Heeger, 2009).

Lastly, one may want to quantify the representational strength in a brain area. In cases where a feature like orientation or color is considered, a set of 'steerable filters' with different preferences are employed in IEM analysis (**Figure 2-4b**). The estimated responses of these filters are further circularly shifted to align with the same center (Ester et al., 2013, 2015). The representational strength can be inferenced from the distribution of circular shifts, which is approximately von Mises distributed in the case of 'steerable filters'. The amplitude and dispersion of the circular shift distribution can be estimated by fitting it with a von Mises function (Ester et al., 2013). Notably, the presumption of this approach is that the basis function exhibits a shape similar to the von Mises distribution. In a case where irregularly-shaped basis functions are employed, an alternative approach is needed. For example, in 0, we construct a categorical basis function based on empirical data, which exhibits an irregular shape. To handle this, the encoding basis function is combined with the multivariate pattern analysis to evaluate the color information measure in the brain.

2.3.3. Comparison between Univariate and Multivariate Pattern Analyses

To analyze fMRI data, both univariate analysis and multivariate pattern analysis (MVPA) are commonly used. However, they sometimes provide different, even contradictory evidence (Kriegeskorte et al., 2006; Riggall and Postle, 2012). In this section, I discuss four main differences between univariate analysis and MVPA, followed by brief discussions on reasons for choosing the analysis implemented in this thesis.

First of all, univariate and multivariate analyses are embedded in different underlying statistical philosophies (Kriegeskorte and Bandettini, 2007; Hebart and Baker, 2017). The conventional univariate analysis focuses on estimating the level of brain activation in response to external stimuli within a standard statistical framework (Worsley et al., 1992; Friston K. J. et al., 1994). In contrast, MVPA targets the amount of information held in multi-voxel activity patterns within an information-based framework (Haxby et al., 2001; Cox and Savoy, 2003; Haynes and Rees, 2006). The former is an activation-based approach and can be seen as assessing the level at which brain regions are engaged or involved in a specific neural function (Jimura and Poldrack, 2012; Hebart and Baker, 2017). In contrast, the latter is an information-based approach, asking how much information is encoded in spatial patterns that could differentiate between experimental conditions (Kriegeskorte and Bandettini, 2007). To summarize,

univariate analysis is activation-based and estimates neural engagement levels, while MVPA is information-based and examines neural representational content (Mur et al., 2009).

They also differ as their respective names suggest: to test the effect at a voxel, univariate analysis examines solely this single voxel, while MVPA jointly examines an ensemble of voxels within the voxel-centered spherical volume (Hebart and Baker, 2017). The former conventionally compares brain activity of different conditions in the voxel (every condition is represented by a one-dimensional brain activity), whereas the latter takes the relation between multiple neighboring voxels into account and contrasts the multi-voxel activity patterns of separate conditions (every condition is represented by a multi-dimensional activation pattern; Haynes and Rees, 2006).

Notably, MVPA takes the activity pattern of multiple voxels into consideration, which includes both activation and deactivation level of single voxels, as well as activity change between voxels (Mur et al., 2009). Different experimental conditions can be represented by different non-uniform spatial patterns in MVPA. In contrast, in univariate analysis, neural responses of neighboring voxels are assumed to be roughly uniform, which justifies spatial smoothing or averaging within the ROI. Considering that voxels within a region respond with opposite activation level to a stimulus, the overall response can cancel out, and thus the engagement of this region is undetected by univariate analysis. In short, the distinction between uniform and non-uniform response patterns of neighboring voxels is another notable difference (Hebart and Baker, 2017).

Univariate analysis can test the statistical difference between a main and a control condition, such as distinguishing between memorizing a digit and seeing an arrow, in order to evaluate the engagement of a brain region in a cognitive process. In this analysis, the effect within a single voxel is typically estimated by a one-dimensional comparison between two conditions. In the case of significantly higher activity in the experimental condition than in the control condition, a positive effect is considered in the voxel. MVPA, in contrast, focuses on a comparison between different variations of the experimental feature (feature-specific), such as memorizing different orientations (Hebart and Baker, 2017). Whether the multi-voxel pattern of one condition is higher or lower than other conditions is of little importance in MVPA. Instead, the degree to which activity patterns of these conditions differ from each other is assessed to reveal the representational level of the feature. In short, while univariate analysis is a directional analysis that examines which condition has significantly higher brain signal than the other,

MVPA tests the level of differentiation between different conditions, and cares little about the direction of the difference.

These and other potential differences can lead to the identification of different functional mechanisms using univariate and multivariate analysis (Riggall and Postle, 2012). Some voxels are identified by univariate analysis but not by multivariate decoding, some are identified by both approaches, while some are marked only by MVPA (Kriegeskorte et al., 2006; Riggall and Postle, 2012). One should choose the appropriate analysis approach based on the research question. In order to examine and understand the specific function of a brain region (a common goal in cognitive neuroscience research), one can design a task to test whether feature-specific content is held in the local area. For example, to investigate whether a brain region stores verbal working memory, one can estimate stimulus-specific content by performing MVPA on brain data acquired during a working memory task. By testing the degree to which conditions (e.g. different verbal stimuli) can be distinguished, MVPA identifies brain regions storing content-specific information. In contrast, by applying univariate analysis, one estimates the engagement level, which can also refer to an indirect participation in memory.

Furthermore, in our studies, the goal of carrying out MVPA is not to construct a predictive model with best decoding performance, but to examine the brain function, more specifically the neural mechanism of working memory. Therefore, instead of performing classifier-based decoding (Cox and Savoy, 2003), we directly estimate the amount of explained multivariate variance, the pattern distinctness D, by using cvMANOVA (Allefeld and Haynes, 2014). Compared to classifier-based MVPA, this approach can estimate standardized effect sizes, provide more explicit assumptions, display equivalent or above sensitivity, as well as avoid some conceptual confusions (Hebart and Baker, 2017). Thus, cvMANOVA better serves our purpose of studying neural mechanisms of working memory and is therefore utilized as the main MVAP method in this thesis.

Chapter 3 Study I: Decoding Verbal Working Memory Representation of Chinese Characters

Brief Summary of This Empirical Study

This study aims to discover how verbal working memory content is retained in the human brain. We designed an experiment to test the short-term memorization of well-known Chinese characters by native speakers. Chinese symbols were uniquely chosen because their simple pronunciation and complex visual appearance heavily facilitate verbal coding. Searchlight-based multivariate pattern analysis was utilized to identify stimulus-selective information present in fMRI signals over the delay period. Three regions were found to carry content-specific information, but the early visual cortex (EVC) allowed for decoding of comparable amount of information for cued and uncued stimuli and was thus more likely to be involved in the perception than the memorization process. Broca's area and left premotor cortex held (1) no significant information about characters not cued for memorization, (2) significantly more information in the left than the right hemisphere and (3) little or no information about memorized complex visual patterns which are hard to verbalize, and thus pointing towards a high specificity. Our findings have provided evidence of verbal WM content stored in language-related areas, consistent with distributed accounts of working memory storage. These active representations of memorized contents in Broca's area and the premotor cortex might constitute the neural substrate of the rehearsal process as part of the "phonological loop" postulated by Baddeley.

3.1 Introduction

The question of how working memory (WM) is maintained in the human brain has received extensive interest for many years. A number of previous neuroimaging studies used multivariate pattern analysis (MVPA) to investigate content-specific WM storage of a wide range of information types. Not only storage of low-level sensory contents including visual features such

as orientation (Harrison and Tong, 2009; Ester et al., 2015) and color information (Serences et al., 2009), motion contents (Riggall and Postle, 2012; Emrich et al., 2013), auditory information (single tones or sounds without any semantic meanings; Linke and Cusack, 2015; Kumar et al., 2016) and tactile patterns (Schmidt and Blankenburg, 2018), have been investigated in the last decade, but also maintenance of spatial locations (Jerde et al., 2012) and object information (Lee et al., 2013).. However, there is so far no direct evidence of verbal WM storage. The primary motivation of this study is to examine the basic neural substrates for the maintenance of verbal WM content. Previously, an EEG study decoded contents of roman letters 'L' or 'T' from the human prefrontal cortex, but the letters were considered visual WM due to the missing semantic meanings and simple visual appearances (Polanía et al., 2011). Some MVPA studies employed English words and pseudowords as stimuli, but decoded the contrast between domains instead of the stimulus-specific content (Lewis-Peacock et al., 2012; Yue et al., 2018). This is the first study investigating the content-specific storage of verbal information using MVPA, which adds the missing piece of evidence for understanding WM storage of the whole range of information formats.

Furthermore, this study can provide important evidence to address an ongoing debate between the 'distributed' storage model and the traditional 'centralized' model of WM. The traditional WM model suggests that specialized systems centralized in the lateral prefrontal cortex (lPFC) serve WM storage and manipulation functions (Goldman-Rakic, 1995). However, this centralization view has been challenged by multiple neuroimaging studies. Researchers decoded stimulus-specific WM contents in different posterior brain areas such as the visual cortex (Harrison and Tong, 2009; Riggall and Postle, 2012; Emrich et al., 2013), auditory cortex (Linke and Cusack, 2015; Kumar et al., 2016), hMT+ (Emrich et al., 2013; Christophel and Haynes, 2014a), the frontal eye fields (FEF; Jerde et al., 2012) and posterior parietal cortex (Christophel et al., 2012; Jerde et al., 2012; Christophel and Haynes, 2014a), depending partially on the stimulus form (more see review from Christophel et al., 2017). Therefore, an alternative 'distributed' model was proposed, arguing for a coordinated recruitment of distributed regions involved in sensory-, representation- or action-related processes (Fuster, 1995; Postle, 2006; Zimmer, 2008a; Christophel et al., 2017).

Notably, a few recent studies found content-specific WM storage of objects (Lee et al., 2013), orientation (Ester et al., 2015) and auditory stimuli (Kumar et al., 2016) in both lateral prefrontal and posterior sensory regions. These findings raised the question of whether the double

representations of memorized contents are redundant. A recent review argues for a labor division between the prefrontal and sensory regions such that, while prefrontal regions represent abstract, semantic or verbal information, sensory regions maintain non-verbal sensory details of stimuli (Christophel et al., 2017). This study examines the whole brain to map brain regions representing delay-period verbal contents, and thus help differentiate between the existing conflicting theories.

Aiming at reaching these goals, we designed a match-to-sample task over an extended delay of 10 s and recruited 30 Chinese native speakers as participants. Importantly, well-known simplified Chinese characters were employed as stimuli. With monosyllabic pronunciation, semantic meaning, but complex visual appearance, visually presented Chinese characters strongly encourage verbal memorization (Zhang and Simon, 1985; Hue and Erickson, 1988). Thus, we chose Chinese characters and not roman letters or words to study verbal working memory. A retro-cue paradigm was employed in order to disentangle the mnemonic from the perceptual brain activity. We measured BOLD activity throughout the brain using fMRI while subjects memorized well-known simplified Chinese characters over an extended delay. Then the entire human brain was probed for activity patterns representing the individual characters using a variant of MVPA: cross-validated MANOVA (cvMANOVA; Allefeld and Haynes, 2014) as well as a searchlight approach (Kriegeskorte et al., 2006).

3.2 Methods

3.2.1. Participants

Thirty healthy right-handed native speakers of simplified Chinese who have been raised up in mainland China and aged between 18 and 35, were recruited in Berlin to participate in the fMRI experiment. However, two participants were excluded from fMRI analyses due to their poor behavioral performance (**Figure 3-3**). The final sample included 28 subjects (15 males and 13 females; age 27.25 ± 0.78 years old). These subjects had normal or corrected-to-normal vision and satisfied all requirements of MRI experiments (e.g. no metal implants). We chose this sample size based on previous lab experiences with content-specific visual working memory studies (Christophel et al., 2012; Christophel and Haynes, 2014a), and increased the subject

number considerably. Subjects gave informed consent and the study was approved by the local ethics committee in the psychology department of Humboldt University in Berlin.

3.2.2. Stimuli

Aiming at a dominant verbal memorization strategy, a set of simplified Chinese characters with high familiarity among native Chinese speakers and comparable visual complexity were employed as stimuli in this study. Simplified written Chinese characters have been officially used in mainland China and Singapore since the 1950s, among which about 7,000 are in general use. Different from a Roman letter or word, each Chinese character comprises only one syllable, has semantic content, and consists of multiple basic strokes, which strongly faciliates the verbal and acoustical memorization even when presented visually (Zhang and Simon, 1985; Hue and Erickson, 1988). Therefore, in this study we chose Chinese characters and not roman letters or words as stimuli.

From all simplied Chinese symbols, we only drew stimuli from the 'List of Frequently Used Characters in Modern Chinese' (Ministry of Education of the People's Republic of China, 1988), which contains the 3,500 most frequently used Chinese characters containing between 1 and 23 strokes. This ensures high familiarity among subjects and therefore enhances the chance of memorizing verbally (Hue and Erickson, 1988). The number of strokes of which a given character is composed is often regarded as a way to measure its visual complexity (Zimmer, 2008a). Here, we chose characters with 12 strokes, because 12-stroke characters are sufficiently complex and are the most frequently appearing, thus providing a sufficient number of characters in the 1988 'List'. To make sure every stimulus we chose can be well verbalized, an additional rating test was completed by five native Chinese speakers. They were asked to evaluate how well they knew each 12-stroke character from the list and its pronunciation (in total 320 characters with 12 strokes from the list). Only characters, which were rated as 'exact recognition' by all five native speakers, were selected for the stimulus pool. Furthermore, Chinese characters with symmetric structure (such as 晶 which means crystal) were left out due to their comparatively low visual complexity. As a result, 240 well-known simplified Chinese characters with 12 strokes were selected for the stimulus pool for this study. Images for all stimuli were taken from a database provided by the Mojikyo institute (http://www.mojikyo.org/), with the same size and font.

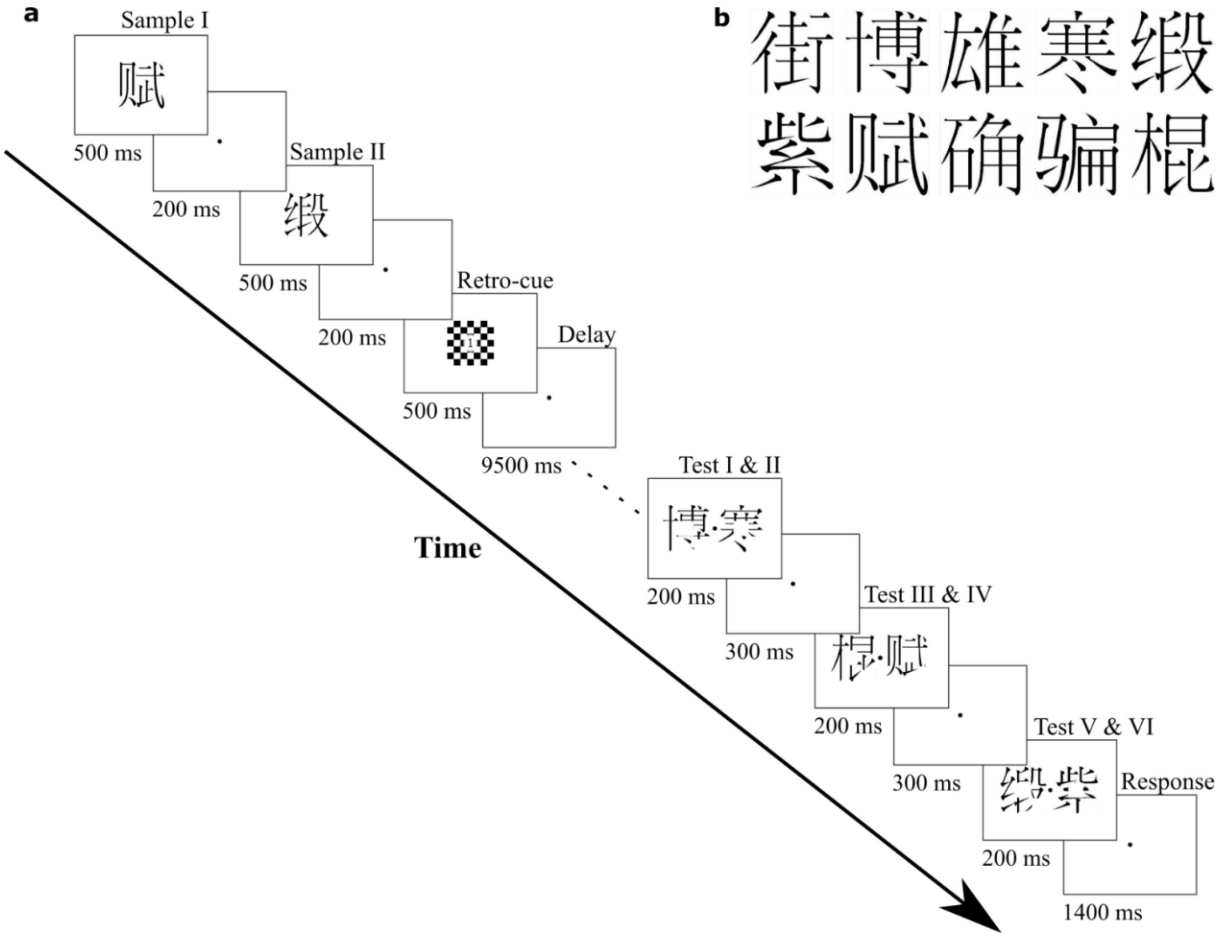

Figure 3-1 a) A trial illustration of the retro-cue based match-to-sample task over an extended delay period of 10 s. In each trial subjects were presented sequentially two sample stimuli. This was followed by a retro-cue ('1' or '2') on a background of a black and white checkerboard, indicating that either the first or the second sample stimulus should be memorized (cued stimulus & uncued stimulus). Then a blank screen (with only the fixation point) was displayed, resulting in an overall retention delay of 10 s (analysis time window). After the delay period, six test stimuli were presented in three sequential screens with two on each screen. Subjects were asked to choose the cued stimulus by pressing the corresponding button. Test stimuli were partly occluded at random positions in each trial to prevent subjects from remembering only parts of the characters. **b)** An individual stimulus set of ten simplified Chinese characters was generated for each subject. Characters belonging to the same set possess different pronunciations and comparable visual complexity. Pronunciations of the illustrated sample set are 'jie, bo, xiong, han, duan; zi, fu, que, pian, gun' (monosyllabic, in pinyin).

For each participant, a different individual sample set was generated with 10 Chinese characters (**Figure 3-1b**) drawn from the stimulus pool. The limited number of sample stimuli per subject

was to allow for the subsequent fMRI multi-voxel pattern analysis. Individual sample sets were generated based on several criteria. Characters belonging to the same sample set possess different pronunciations (neither common consonants nor common vowels), low pixel-by-pixel correlation (Pearson correlation ≤ 0.1), and similar proportions of black pixels (difference ≤ 10%).

3.2.3. Experimental Paradigm

This is a retro-cue-based match-to-sample task over an extended delay period of 10 s (**Figure 3-1a**). A trial began with the sequential presentation of two sample stimuli. Each sample stimulus was shown for 500 ms, followed by a fixation period of 200 ms. Then, a retro-cue ('1' or '2') was presented on a black and white checkerboard background for 500 ms. The retro-cue instructed subjects which of the two sample stimuli to remember for the trial (cued & uncued sample; Sperling, 1960). In fMRI analysis, this retro-cue procedure serves to disentangle the mnemonic from the perceptual brain activity. Then, a blank screen with only the fixation point was displayed for 9.5 s, resulting in an overall delay of 10 s. This 10 s delay period is the time window for our fMRI analysis on WM process. Afterwards, a sequence of three test screens was shown, each for 200 ms, separated by 300 ms. Each test screen contained two test characters. Subjects were required to find the cued sample character among six test characters. This difficult variant of a match-to-sample task required subjects to identify the memorized item out of a larger set of stimuli (chance level: 16.67%) to minimize the ceiling effect. After the offset of the third test screen, they had 1400 ms to respond by pressing the corresponding button. Participants were required to fixate on the fixation point in the middle of the screen throughout the experiment.

All test stimuli were partly occluded to discourage subjects from memorizing only a small part of the character. Randomly, two out of nine patches that covered the whole stimulus space were occluded for all test characters in each trial. The trial duration was 14 s, followed by a varying inter-trial interval of 2 to 8 s (on average: 4.8 s). Every subject completed one fMRI session of 4 runs, with 50 trials per run. Every Chinese character in the sample set was presented five times per run and 20 times in total as the cued sample. The trial order was randomized. Stimuli were presented via a projector during scanning. The experiment was programmed using Matlab (Mathworks, Natick, MA) and Cogent 2000 (http://www.vislab.ucl.ac.uk/cogent_2000.php).

3.2.4. Data Acquisition

All fMRI data were collected on a Siemens 3 Tesla Trio scanner at the Berlin Center for Advanced Neuroimaging (BCAN). Both high-resolution structural MRI data (T1-weighted MPPRAGE: 192 sagittal slices; TR = 1900 ms; TE = 2.52 ms; flip angle = 9°; FOV = 256 mm) and functional BOLD images (T2*-weighted gradient-echo EPI: 32 contiguous slices; whole neocortex; TR = 2 s; TE = 30 ms; voxel size = 3x3x3 mm; matrix size = 64 × 64 × 32; slice gap = 0.6 mm; descending order; flip angle = 90°; FOV = 192 mm) were acquired for each subject. In every experimental run, 473 functional images and altogether 1892 functional scans were collected per participant. The trial onset was time-locked to the acquisition onset of an fMRI image to reduce temporal variability in the data analysis.

Behavioral responses were collected via a pair of MRI-compatible button boxes with 2 × 4 buttons (using the first three buttons on both sides). Furthermore, subjects were asked to fill out a questionnaire after finishing the experiment, which evaluates their strategies for memorizing Chinese characters. Various strategies were listed in the questionnaire, including memorizing the stimulus as acoustic, semantic, visual, or spatial information, as well as through a related emotion, action, touch, smell or taste.

3.2.5. fMRI Analysis

This section describes the fMRI analysis conducted to estimate brain regions that carry content-specific WM of Chinese characters during the delay period. The key part is a variant of MVPA: cvMANOVA (Allefeld and Haynes, 2014) combined with the searchlight procedure (Kriegeskorte et al., 2006). Statistical fMRI analysis was performed using SPM8 (Friston K. J. et al., 1994) and cvMANOVA (Allefeld and Haynes, 2014).

3.2.5.1. Preprocessing

The acquired fMRI data were first converted from DICOM to NIfTI format and then spatially realigned and resliced to correct head movements. No other preprocessing (e.g. normalizing and smoothing) was conducted in order to maintain the fine spatiotemporal properties of the BOLD activity and therefore maximize the sensitivity of multivariate pattern analysis (see section 2.2 for more about fMRI preprocessing).

3.2.5.2. Univariate Model

The preprocessed brain data was fitted by a generalized linear model (GLM) to estimate the event-related BOLD activity over the delay (see section 2.3.1 for more about univariate analysis). In response to each cued sample, five finite-impulse response (FIR) regressors were used to represent the five scans during the 10 s delay period (TR = 2s). For 10 memorized characters and four runs, there were 200 regressors altogether (10 samples × 5 scans × 4 runs) for each subject. The estimated results can be further used for the following multivariate pattern analysis.

3.2.5.3. Searchlight-based MVPA

A variant of MVPA, the cvMANOVA approach, was used to identify brain activity patterns that could differentiate between experimental conditions (see section 2.3.2.3 for more about computation of cvMANOVA). It assesses the amount of multivariate between-class covariance explained by the effect relative to the within-class error covariance in a leave-one-session-out cross-validation paradigm (Allefeld and Haynes, 2014). A contrast matrix was built to target the main effect of memorized Chinese characters (**Figure 3-2**). The 10 s delay period was modeled by 5 finite impulse response (FIR) regressors in response to the memorized stimulus. For a given FIR bin, 10 regressors (ten sample Chinese characters) were compared in neighboring pairs resulting in 9 contrasts. Over all five FIR bins, there were 50 regressors and 45 contrasts. The resulting unbiased estimate of the multivariate covariance explained by the main effect in units of error covariance is called *pattern distinctness D* (Allefeld and Haynes, 2014). Furthermore, cvMANOVA was combined with a searchlight analysis to map the entire

brain (Kriegeskorte et al., 2006). For each voxel in the brain, we drew all the voxels within the voxel-centered spherical volume with a radius of 5 voxels for joint analysis. This resulted in a statistical parametric map of pattern distinctness D throughout the brain SPM$\{D\}$.

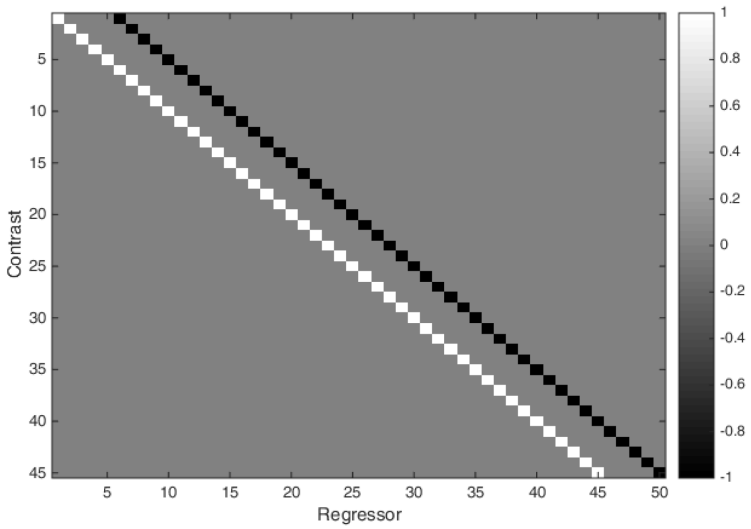

Figure 3-2 The transposed contrast matrix C' between ten memorized Chinese characters in a WM task with a 10 s delay, each modeled by five FIR regressors. The pairwise (1,-1) contrast between neighboring conditions leads to 9 contrasts for a given FIR bin, and in total 45 contrasts for five FIR bins.

3.2.5.4. Post-processing

The estimated SPM$\{D\}$ map was normalized using the co-registered, segmented anatomical image as the deformation field, and then smoothed with a Gaussian kernel with a FWHM of 5 mm (see section 2.2 for more details).

3.2.5.5. Group-level Statistics

The post-processed SPM$\{D\}$ maps of 28 subjects were statistically tested using a one-sample one-sided t-test against the zero baseline, in order to infer the group-level effect. A Bonferroni correction procedure was employed to correct the family-wise error (FWE) caused by group-level analysis on multiple voxels in parallel (Bonferroni, 1936; Dunn, 1961).

3.3 Results

3.3.1. Behavioral Results

Native Chinese speakers performed well on the working memory task of Chinese characters in general (**Figure 3-3**). On average, the 30 subjects achieved a response accuracy of 85.97% in the delayed match-to-sample task. We excluded two subjects from further fMRI analysis because their performance lay more than two standard deviations below the mean group accuracy of 68.4%. The remaining 28 subjects reached an accuracy of 87.64% on average, with a standard error of mean (SEM) of 1.17%. The average reaction time to respond was 735 ms across 28 subjects, with a SEM of 26 ms.

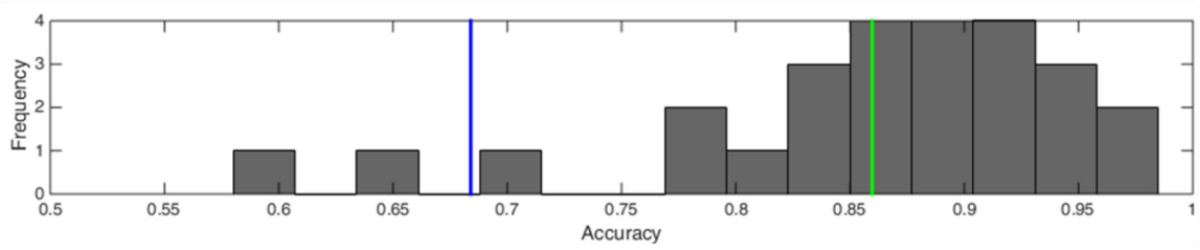

Figure 3-3 Distribution of behavioral performance of 30 Chinese native speakers. Two subjects were excluded due to their poor behavioral performance that lay more than two standard deviations below the group average. The remaining 28 subjects achieved an average accuracy of 87.64% with the SEM of 1.17%. Green line: group average performance across all 30 subjects (85.97%); blue line: two standard deviations below the 30 subject group average (68.4%).

3.3.2. Questionnaire Results

Subjects were asked to evaluate how accurately each statement in the questionnaire matched their strategies for memorizing the cued stimulus over the delay period, by using a score ranging from 0 to 7 (0 means 'not apply at all', while 7 means 'applies fully throughout the experiment'). Rating scores were compiled across 28 subjects, and histograms for all 12 statements are illustrated in **Figure 3-4**. While these statements were shown in random order to individual

subjects, they are ordered here in descending median rating score. The top two statements: 'memorized as how it sounded' and 'memorized as what it meant' referring to working memory in acoustic and semantic form, both belong to the broad concept of verbal working memory. Participants also memorized the target character visually as how it looked, but much less frequently than verbal memorization.

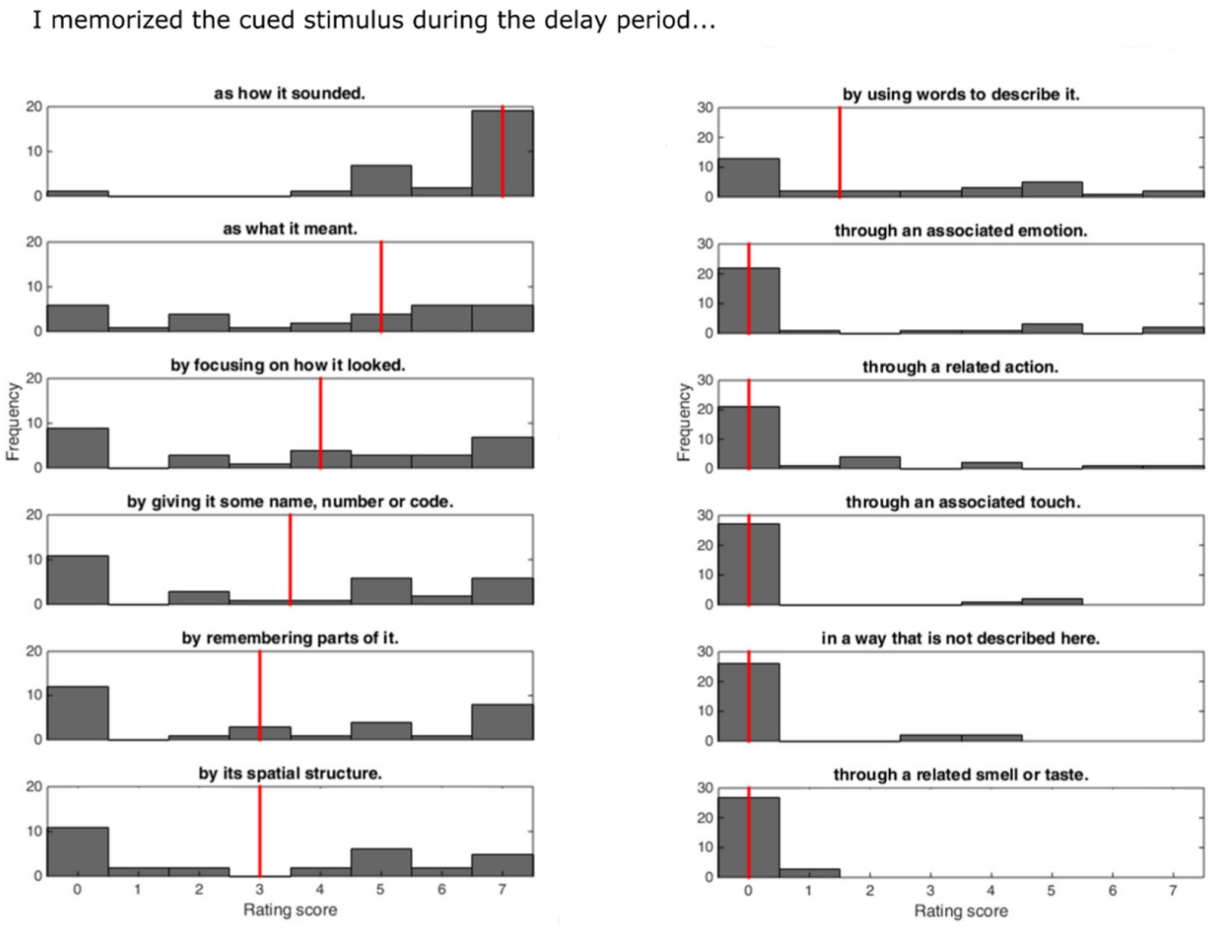

Figure 3-4 Histograms of the questionnaire results. After the fMRI experiment subjects were asked to evaluate how accurately each statement described their strategy to accomplish the working memory task using rating scores from 0 to 7 (0: applies not at all; 7: applies fully throughout the task). Statements were presented in random sequence to individual subjects, but ordered here in descending median rating score. Red line: the median rating score among 28 participants.

3.3.3. fMRI Results

By employing the searchlight-based cvMANOVA approach, we calculated throughout the brain the unbiased estimate of the explained multivariate covariance between memorized Chinese characters, called pattern distinctness D. A three-dimensional parametric map SPM{D}, which reflects the information measure of memorized Chinese characters on voxel level over the whole brain, was obtained. The group-level statistics across 28 subjects revealed brain clusters that held significant stimulus-specific information over the 10 s delay period. In this section, four questions are addressed to closely examine the decoding results.

(i) Which Brain Regions Retain Significant Delay-period Information about Memorized Chinese Characters?

The estimated three-dimensional statistical map of multivariate pattern distinctness D throughout the brain was tested across 28 subjects using a cluster-level corrected one-sided one-sample t-test. Three brain regions were identified as carrying significant information about cued Chinese symbols during the delay period ($P_{FWE} < 0.05$, cluster-level corrected with cluster-defining threshold of $P < 0.001$; **Figure 3-5** & **Table 3-1**).

One brain region covers the pars orbitalis and pars triangularis of the left inferior frontal gyrus (Brodmann area 45, 46 & 47; cluster-level corrected $P_{FWE} = 0.038$), overlapping with the anterior part of what is generally considered as Broca's area (Pulvermüller and Fadiga, 2010). Another cluster maintaining significant content-specific information was found in left premotor cortex within the frontal lobe (Brodmann area 6, 8 & 9; cluster-level corrected $P_{FWE} < 0.001$). More specifically, this cluster is located anterior to the primary motor cortex and covers a major part of the precentral gyrus and the posterior part of the middle frontal gyrus in the left hemisphere. We found the last significant cluster in the early visual cortex (often considered to include V1, V2 and V3) in both left (Brodmann area 17 & 18; cluster-level corrected $P_{FWE} < 0.001$) and right hemispheres (Brodmann area 17, 18 & 19; cluster-level corrected $P_{FWE} = 0.004$). This cluster is located in the posterior end of the occipital cortex. For simplification, in this thesis, these three clusters are addressed as anterior Broca's area (aBA), left premotor cortex (lPMC), and early visual cortex (EVC).

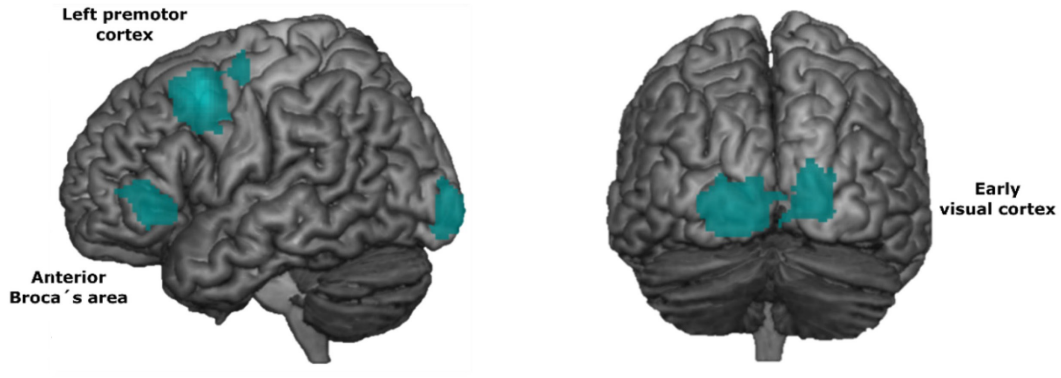

Figure 3-5 Three brain areas carried significant content-specific information about cued Chinese characters during the delay period (one-sided one-sample t-test with $P_{FWE} < 0.05$, cluster-level corrected with cluster-defining threshold of $P < 0.001$; N=28).

Brain region	Cluster-level			Peak-level				
	Hemis-phere	Brodmann area	$P_{(FWE)}$	MNI coordinate X	Y	z	T	$P_{(uncorr)}$
Anterior Broca's area	left	45,46,47	0.038	-56	34	-4	5.24	< 0.001
				-58	26	2	3.95	< 0.001
				-54	40	8	3.93	< 0.001
Left Premotor cortex	left	6,8,9	< 0.001	-46	10	48	7.33	< 0.001
				-36	-8	60	4.23	< 0.001
Early visual cortex	left	17,18	< 0.001	-10	-92	4	5.00	< 0.001
				-16	-98	-2	4.56	< 0.001
				-22	-98	6	4.50	< 0.001
	right	17,18,19	0.004	16	-100	10	4.62	< 0.001
				16	-100	-2	4.57	< 0.001
				24	-96	0	4.32	< 0.001

Table 3-1 Brain regions that held significant content-specific information of cued Chinese symbols during the delay period (one-sample one-sided t-test with PFWE < 0.05, cluster-level corrected with cluster-defining threshold of P < 0.001; N=28).

(ii) Is the Identified Information Specific to Working Memory or Perception?

The retro-cue-based paradigm was employed in order to disentangle the brain activity of memorization from perception. Two Chinese characters were presented while one was cued to be memorized. While the cued sample elicits brain activity in response to perceiving and memorizing the cued stimulus in sequential order, the uncued sample only engenders the brain signal in response to perceiving the visual sample. Because the hemodynamic response is not instant but delayed and extended over several seconds, the BOLD activity during the delay

period can be caused by perceptual signal, mnemonic signal, or a combination of both. One good way to differentiate between these two processes is to compare the brain signal of the cued sample with that of the uncued sample. A brain region that retains significant information of the cued sample but little of the uncued sample, and exhibits a significant difference between the two conditions, is specific to working memory, not perception. In contrast, a brain area holding comparable amount of information about the cued and the uncued sample is likely to take part in perceiving the Chinese character but less likely to participate in memorizing it.

This was tested on all three brain regions across 28 subjects, using the average pattern distinctness of all cluster-peaks (**Figure 3-6**). We found that the anterior Broca's area and left promotor cortex carried little information for uncued stimuli (two-sided one-sample t-test; aBA: $P = 0.505$; lPMC: $P = 0.571$) and significantly more information for cued sample (two-sided paired t-test; aBA: $P = 0.004$; lPMC: $P < 0.001$), thus qualifying as working memory stores. The early visual cortex, however, showed a significant level of information for uncued stimuli (two-sided one-sample t-test; $P < 0.001$) and a similar amount of information for cued and uncued symbols (two-sided paired t-test; $P = 0.543$), and is thus unlikely to be involved in the retention process.

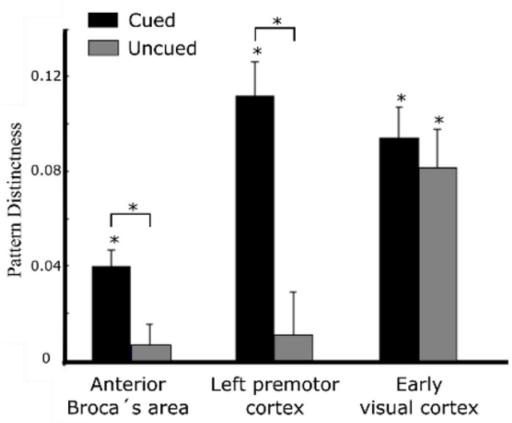

Figure 3-6 Comparison between representations of memorized stimulus (cued sample) and not-cued-to-be-memorized stimulus (uncued sample) in all three regions. Regions encoding a comparable level of information for cued and uncued samples were assumed to be not specific to retention and thus excluded from further analyses. Error bars represent between-subjects SEM; * above bars refers to two-sided one-sample t-test with $P < 0.05$; * above brackets indicates two-sided paired t-test; $N = 28$.

(iii) Is Stimulus-specific Working Memory Left Lateralized?

Both clusters we found to retain memory contents of Chinese characters during the delay period are in the left hemisphere. We estimated the pattern distinctness D of the corresponding cluster-peaks in the right hemisphere, averaged D across peaks within the cluster, and statistically tested the difference between two hemispheres across 28 subjects (**Figure 3-7**). No significant information was found in the right brain regions mirror to the anterior Broca's area or the left premotor cortex (two-sided one-sample t-test; aBA: $P = 0.098$; lPMC: $P = 0.436$). Furthermore, by conducting a two-sided paired t-test between the left and the right hemisphere, we found significantly more information retained in the left than in the right hemisphere (aBA: $P = 0.046$; lPMC: $P < 0.001$).

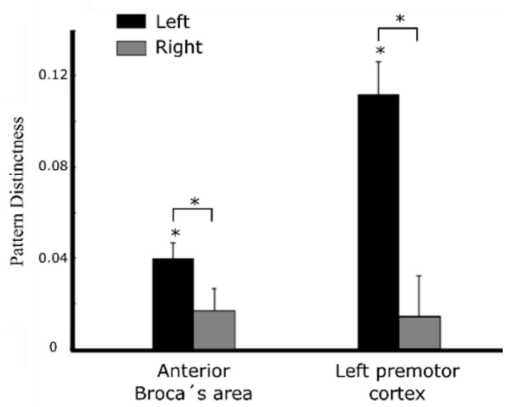

Figure 3-7 Comparison of the information measure between the left and the right hemisphere in brain regions that exclusively represent the mnemonic content. Error bars represent between-subjects SEM; * above bars refers to two-sided one-sample t-test with $P < 0.05$; * above brackets indicates two-sided paired t-test with $P < 0.05$; $N = 28$.

(iv) Are Identified WM Stores Specific to Verbal Material?

In this section we contrasted our study on Chinese characters with non-verbal conditions from previous studies, to clarify whether the identified working memory stores in Broca's area and the left premotor cortex retain specifically verbal material. For this reason, we selected two previous working memory studies investigating the retention of visual patterns for contrast. Both studies employed the retro-cue-based paradigm and match-to-sample task. Furthermore,

exactly the same searchlight-based cvMANOVA approach was utilized to estimate the information measure over the 10 s delay period throughout the brain. Due to employment of similar experimental design and analytical procedure, this comparison of information estimates is valid (Hebart and Baker, 2017). One study investigated the working memory of complex color patterns (Christophel et al., 2012), while the other tested the maintenance of complex motion patterns (Christophel and Haynes, 2014a). Both Stimuli (color and motion patterns) were visually complex enough, with randomized pattern structure, that they were hard to describe by words.

This study is statistically compared with each of the two previous studies. The sample size was 28 in this study, and 17 in both previous studies. Therefore, we conducted a two–sided two-sample t-test on parametric maps of pattern distinctness in brain areas holding working memory contents of Chinese characters (analysis based on the average pattern distinctness of all peaks in the respective cluster; **Figure 3-8**). Both the anterior Broca's area and left premotor cortex retained significantly more information for Chinese characters than for complex color patterns (two-sided two-sample t-test; aBA: $P < 0.001$ and lPMC: $P = 0.003$), and complex motion patterns (two-sided two-sample t-test; aBA: $P = 0.001$ and lPMC: $P = 0.003$) during the 10 s delay period.

Furthermore, Broca's area and the left premotor cortex contain little information about complex visual patterns of color or motion during the delay period (two-sided one-sample t-test; complex color patterns in aBA: $P = 0.494$ and in lPMC: $P < 0.001$; complex motion patterns in aBA: $P = 0.804$ and in lPMC: $P = 0.082$). The only exception was that the left premotor cortex contained significant information about complex color patterns, although significantly less than information about Chinese characters. In summary,, Broca's area and the left premotor cortex serve as working memory stores specific to verbal rather than visual material.

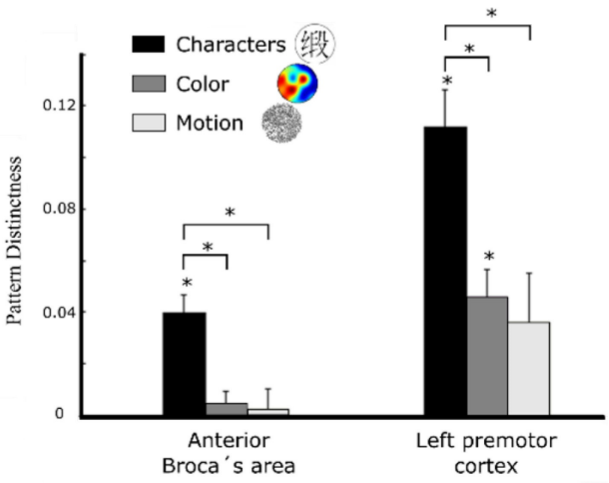

Figure 3-8 Comparison of information content in the current study with two previous studies on WM storage of complex visual patterns that were hard to verbalize (complex color patterns & complex motion patterns; N= 28 + 17 + 17; see Christophel et al., 2012; Christophel and Haynes, 2014a). Error bars represent between-subjects SEM; * above bars refers to two-sided one-sample t-test with $P < 0.05$; * above brackets indicates two-sided two-sample t-test with $P < 0.05$; N = 28, 17 and 17 in WM study on Chinese characters, complex color patterns and complex motion patterns respectively.

3.4 Discussion

The present study employed a brain mapping approach to estimate brain regions maintaining content-specific information on Chinese characters during a 10 s delay period across 28 native Chinese speakers. We found three brain regions holding significant information about memorized Chinese characters (**Figure 3-5**). But only two regions, anterior Broca's area and left premotor cortex retain working memory contents, while the other area, the early visual cortex, carries a comparable information measure of cued and uncued stimulus, and is thus unlikely to be involved in working memory (**Figure 3-6**).

We found significantly more information in Broca's area and the premotor cortex in the left hemisphere as compared to their right-hemispheric counterparts (**Figure 3-7**). The predominance of left-hemispheric areas in right-handed participants is a hallmark of language processing (Pulvermüller and Fadiga, 2010), Broca's area is known as a key area for language production, and lesions of this area severely impairs speech production (Broca, 1861). Its

involvement in language processing such as grammar processing has also been found in multiple fMRI studies (Just et al., 1996; Kinno et al., 2008). The premotor cortex has been shown to be important for the planning, the preparation, the selection and the initiation of movement (Wise, 1985). Stimulation of the left premotor cortex, however, also induces transient speech disturbances (Duffau et al., 2003). Additionally, recent TMS and fMRI evidence suggest its involvement in silent articulation and language comprehension (Iacoboni, 2008; Schomers et al., 2015).

Previous evidence exists supporting the participation of Broca's area and the left premotor cortex in working memory. Lesion work suggested that damage to the left premotor cortex can lead to severe impairment in rehearsal and deficits in verbal short-term memory (Vallar et al., 1997). In addition, previous univariate fMRI research found that the premotor cortex and ventral-lateral prefrontal cortex showed increases in overall delay-period brain activity when subjects memorize visually shown words (Buchsbaum et al., 2005). Furthermore, recent fMRI work showed linearly increased brain activity with a rising verbal rehearsal rate in premotor and inferior frontal areas (Fegen et al., 2015). These previous studies, however, due to their study design and univariate nature, could not demonstrate working memory storage by identifying representations of individual characters.

The anterior Broca's area and left premotor cortex contain significantly more delay-period information for Chinese characters than for complex visual patterns (**Figure 3-8**). Furthermore, they carry little delay-period information for visual stimuli, with the exception that the left premotor cortex stored significant information for complex color patterns. It is unclear whether the limited involvement of the premotor cortex in the storage of color patterns can be attributed to its general involvement in visual memory (e.g. active refreshment of visual stimuli during the delay) or the occasional use of verbalization strategies for complex color patterns. Participants further reported a dominance of verbal (incl. acoustic and semantic) strategies in memorizing cued Chinese characters during the delay period in a post-study questionnaire (**Figure 3-4**). These results suggest that visually displayed characters are memorized verbally, and that Broca's area and left premotor cortex serve as WM stores specific to verbal rather than visual material. Although one cannot completely exclude the possibility that Chinese characters are occasionally memorized visually, it is unlikely that visual WM storage takes place in the Broca's and left premotor cortex.

The two brain regions we found to be verbal working memory stores have large overlaps with an articulatory network from the dorsal stream of language processing. Language processing is considered to be realized through two processing streams: one ventral stream and one dorsal stream (Hickok and Poeppel, 2007), comparable with the dual-stream model of vision (Ungerleider and Mishkin, 1982). This dual-stream model of language processing argues that the ventral stream is mostly bilaterally organized and contributes to the perception as well as recognition of auditory information; while the dorsal stream is strongly left lateralized and plays a role in the interaction between an auditory and a motor system (Hickok and Poeppel, 2007). The articulator network of the dorsal language processing stream is comprised of the posterior inferior frontal gyrus, premotor cortex and anterior insula in the left hemisphere (Hickok and Poeppel, 2007).

From another perspective, an influential cognitive model of WM suggests that the function of verbal working memory storage is accomplished through the tight interplay between two systems: a sensory component and a motor component (Baddeley et al., 1984; Baddeley, 1992). The former captures the phonological input and decays over time, and the latter assists with maintaining the memorized contents via articulatory rehearsal. Although our design cannot discern the exact neural coding schemes used to retain verbal material, based on this Baddeley's 'phonological loop' model and discussions on language processing, one could speculate that Broca's area and the premotor cortex compose the articulatory network which serves the articulatory rehearsal of the verbal working memory function (Jacquemot and Scott, 2006; Hickok and Poeppel, 2007).

Previous fMRI evidence showed that the ventral lateral prefrontal cortex (including Broca's area), premotor cortex (and surrounding motor and sensorimotor cortices) and Sylvian-parietal-temporal area (containing Wernicke's area) display significant delay-period BOLD activity when subjects memorize visually shown words (Buchsbaum et al., 2005). However, in this study, no significant content-specific representation was estimated in posterior brain regions of the left hemisphere, which have been implicated in verbal WM by lesion studies (Warrington and Shallice, 1969; Warrington et al., 1971). This might be a result of representations in these regions failing to cross the threshold of significance, possibly due to power or sensitivity restraints. Further studies could be conducted to address the interplay of the two components of the phonological loop. In summary, language-related regions, including the anterior Broca's area and left premotor cortex, maintain language content in verbal format in WM, possibly

through articulatory rehearsal. This completes to the first research objective of investigating the neural basis of verbal working memory content.

Our findings of verbal WM contents stored in Broca's area and the left premotor cortex display little overlap with the traditionally considered center for working memory storage in lPFC (Goldman-Rakic, 1995). Although both regions lie within the frontal lobe, Broca's area is located inferior and premotor cortex located superior to the lPFC region. In fact, this is the first MVPA study decoding working memory contents from a well-known language region like Broca's area. Thus, our findings argue against the 'centralized' WM model, but provide supporting evidence for the 'distributed' model of working memory storage (Postle, 2006; Zimmer, 2008a; Christophel et al., 2017). This model argues that diverse sensory and non-sensory brain regions covering the neocortex can maintain persistent representations of the content they process, supported by multiple neuroimaging studies using MVPA (for overview see Christophel et al., 2017). To name a few, orientation and color information was decoded from the early visual cortex (Brouwer and Heeger, 2009; Harrison and Tong, 2009), auditory information was decoded from the auditory cortex (Linke and Cusack, 2015; Kumar et al., 2016), motion flow patterns were decoded from hMT+ (Emrich et al., 2013; Christophel and Haynes, 2014a), spatial locations were decoded from the frontal eye field (Jerde et al., 2012), and complex visual patterns were decoded from posterior parietal cortex (Christophel et al., 2012; Christophel and Haynes, 2014a). Our findings provide answers to the second research question and reveals that verbal WM maintenance depends on a distributed network of language-related brain areas, but fail to identify representations in posterior language-specific areas.

Notably, the searchlight-sphere in the current study centered in Broca's area ($peak_{MNI}$ = [-56, 34, -4], radius = 15 mm) overlaps with evidence for putatively 'visual' working memory storage of orientation in the ventrolateral prefrontal cortex ($peak_{MNI}$ = [-37, 30, -2], radius = 8 mm; Ester et al., 2015). In addition, a recent fMRI study revealed content-specific tactile information in the premotor cortex during working memory (Schmidt and Blankenburg, 2018). These pieces of evidence might suggest that verbal working memory plays a role in the maintenance of other kinds of stimuli that have been originally thought of as 'low-level' sensory (see also Spitzer et al., 2014; Vergara et al., 2016), which could be examined in the future work.

This study on verbal material was contrasted with two previous studies on visual material that had comparable experimental design and analysis methods. This can be considered a limitation

of the study, because a direct comparison between verbal and non-verbal (visual) conditions within the same experiment was not provided. In fact, a direct contrast condition was planned but was proven unachievable during the piloting phase. Non-Chinese speakers (German native speakers) were recruited to memorize visually complex Chinese characters, but contrary to expectations, they reported heavy usage of verbal strategies. This suggests that it is difficult to memorize visually presented complex stimuli purely visually with no help of verbalization. A similar result was found in another contrast condition where Chinese native speakers memorized extinct Tangut symbols, which consist of many strokes and are unreadable to Chinese speakers, strongly as verbal information. Further future work could be done to investigate the interplay between the motor component (e.g. Broca's area and left premotor cortex) and the sensory component (e.g. parietal-temporal area) of the phonological loop.

Chapter 4 Study II: Decoding Dual-content Neural Representation of Color Working Memory from the Sensory Cortex

A Brief Summary of this Empirical Study

Traditionally, color working memory is modeled as a continuous representation of hues. However, intuitively, color memorization is also categorical, and recent evidence indicate that color can be memorized as categorical color terms (e.g. ‘yellow’, ‘blue’, ‘green’, etc.). Both 0 and this chapter aim to examine the neural representation of color working memory. In 0, behavioral patterns are examined and a cognitive model is built to test the contribution of the categorical pathway, while in this chapter mainly the neural basis of color working memory is tested.

Subjects performed a pair of delayed and undelayed estimation (DE and UDE) tasks in the fMRI scanner, as well as a pair of color categorization tasks to estimate their color categorization preference. To analyze the fMRI data, we constructed two encoding models: a conventional visual encoding model to characterize the visual representation, and a novel categorical encoding model based on empirical data to characterize the categorical representation of color in the brain. By combining encoding models with a MVPA approach, we identified feature-selective representation of color in three regions of interest (ROIs) that exhibited neural selectivity to color: V1, V4 and VO1. Furthermore, by comparing two encoding models, we estimated the predominant neural representation form. We examined the content-specific color information mainly during working memory (DE task), but also during perception (UDE task), in order to test the interaction effect between encoding models and tasks.

We found significant color information in all three ROIs during working memory. Furthermore, we found that more anterior areas, V4 and VO1, held memorized color information predominantly in categorical form. Additionally, VO1 exhibited a clear interaction effect: its predominant categorical representation was statistically dominating in the DE task. Our

findings implied a gradient of abstraction in the memorized content along the rostral-caudal axis of the brain.

4.1 Introduction

The neural basis of color representation has generated broad interest but remains unclarified. Color vision has been examined by several experimental studies. It was found that color stimuli are first perceived by three types of cone photoreceptors in the retina and then projected to visual cortex via LGN (Dow and Gouras, 1973; Derrington et al., 1984; Kaiser and Boynton, 1996; Solomon and Lennie, 2007). In V1 (the primary visual cortex), a large number of neurons have been found to respond robustly to chromatic stimuli but not to achromatic modulation (Solomon and Lennie, 2007). A majority of neurons in macaque V4 were found to be tuned to chromatic stimuli (Zeki, 1974), and similarly, neuroimaging studies demonstrated human V4 (the fourth visual field map) as a color center with chromatic selectivity (McKeefry and Zeki, 1997; Bartels and Zeki, 2000). It should be noted however, that conflicting evidence also exists, leading to an ongoing debate over the role of V4 in color perception (Heywood et al., 1992; Hadjikhani et al., 1998). Furthermore, evidence showed that an adjacent region to V4, VO1 (the ventral occipital cortex), responds to color stimuli changes (Brewer et al., 2005), and a lesion in this region can cause loss of color-selectivity (Meadows, 1974). In addition to an early lesion, electrophysiology and fMRI univariate approaches, the more recently developed multivariate pattern analysis (MVPA) approach can decode feature-selective representations of color in human brain. For example, in an MVPA-based fMRI study, perceived color information (of eight color samples) was decoded from the human visual cortex, including V1, V4 and VO1 regions (Brouwer and Heeger, 2009). These previous studies demonstrate that brain regions in the visual cortex, in particular, V1, V4 and VO1, exhibit neural selectivity to color.

In contrast to the abundance of studies on color vision, only a small number of fMRI studies were targeted at decoding the WM of color. One of the studies is limited because it utilized a set of only two color samples with small jitters (Serences et al., 2009). Another study employed hard-to-verbalize complex color patterns as stimuli, aiming at decoding visual WM content and

not the pure color representation (Christophel et al., 2012). Thus, these studies could not provide direct evidence of the storage of color content.

Traditional cognitive models of color working memory are based on the assumption of the continuous visual representation of hue (Huttenlocher et al., 2000; Zhang and Luck, 2008). However, this view has been recently challenged. It was alternatively proposed that color working memory is realized by combining categorical representation with continuous visual estimates (Bae et al., 2015). For example, basic color terms like 'blue', 'pink', 'green', 'orange', 'purple', 'yellow', 'red' and 'brown' (Berlin and Kay, 1969) can be utilized to assist in color memorization.

This study mainly aims to 1) identify feature-selective representation of color information in the human visual cortex; 2) test the hypothesis that colors are represented not only visually as continuous estimates, but also as color names or categories in the brain; 3) identify the predominant neural representation form in a brain region. In addition, we are interested in 4) comparing the undelayed visual perception process with the delayed working memory process of colors in the brain and testing the interaction effect between models and tasks; and (5) contrasting the average-based with the individual-based categorical encoding model.

While subjects complete a pair of delayed (Wilken and Ma, 2004; Zhang and Luck, 2008; Bays et al., 2009, 2011; Fougnie et al., 2010; Fougnie and Alvarez, 2011; van den Berg et al., 2012; Bae et al., 2015) and undelayed (Gold et al., 2010; Brady et al., 2013; Bae et al., 2014, 2015; Souza et al., 2014) estimation tasks, we measure their brain activity to study mnemonic or perceptual neural representation of color. Furthermore, subjects are required to perform a pair of categorical tasks, in order to evaluate color categorical preferences (Witzel and Gegenfurtner, 2013; Bae et al., 2015).

To estimate color-selective responses in the brain, a set of basis functions from the inverted encoding model (IEM; see section 2.3.2.4) are utilized, which characterize the selective neural response to color (Engel et al., 1997a; Brouwer and Heeger, 2009). The conversion from the continuous feature space into the basis function space allows content-specific analysis of a large number of color stimuli. Two types of basis functions are employed to model the dual-content neural representation of color. The conventional cosine-shaped basis function is used to characterize the low-level visual representation of color (Brouwer and Heeger, 2009; Ester et al., 2015; Sprague et al., 2016). We construct a novel type of basis function based on empirical

color categorization data, in order to model the categorical neural representation of color. The conventional visual encoding model and the novel categorical encoding model are further contrasted to estimate the predominant representation form in a brain region. In order to quantitatively estimate the feature-specific representation of color, the utilization of the encoding models from IEM is further combined with an MVPA approach, cvMANOVA (Allefeld and Haynes, 2014). The fMRI analysis focuses on three regions of interest (ROIs) that were revealed to exhibit color selectivity: V1, V4 and VO1.

4.2 Methods

4.2.1. Participants

Ten right-handed healthy German native speakers (aged 18-35 years; mean age: 27, SEM ± 1.13; 9 female) with normal or corrected-to-normal vision and no color blindness participated in the study. Each subject completed five sessions of experiments, including three 2-h fMRI sessions with 16 runs (50 trials/run) for the delayed estimation task, one 2-h fMRI session with 14 to 16 runs (50 trials/run) for the undelayed estimation task, and one 90-min behavioral session for categorization tasks. The sample size (the total trial number per task) was chosen based on previous studies using the IEM approach to study WM (Brouwer & Heeger, 2009; Sprague et al. 2016), and was considerably increased. With the same total scanning length, we decided to recruit a small subject number with multiple sessions per subject, instead of a large number of subjects, aiming at minimizing the effect of individual differences in brain shape and size (Cosgrove et al., 2007). This study was granted ethical approval by the local ethics committee and all subjects gave informed consent.

4.2.2. Stimuli

In Commission Internationale de l'Eclairage (CIE) LAB space, a set of 50 color samples with equal spacing in a*b* space (a*center = 0, b*center = 0, radius = 38; **Figure 4-1a**) and a constant lightness (L*=70; **Figure 4-1b**) were generated (Bae et al., 2015). The CIELAB space is a nonlinear conversion of the CIEXYZ space, and is designed to be 'device-independent' and perceptually more uniform (Commission Internationale de l'Eclairage, 1986). In contrast to the

limited sample number in the traditional MVPA-based fMRI studies (8 color samples in Brouwer and Heeger, 2009; 2 color samples in Serences et al., 2009; 4 complex color patterns in Christophel et al., 2012), we employed a large number of 50 color samples, to facilitate the examination of continuous visual representation of color. These 50 color exemplars form a complete circular color wheel (360 degree) with an equal spacing of 7.2 degrees between neighboring hues.

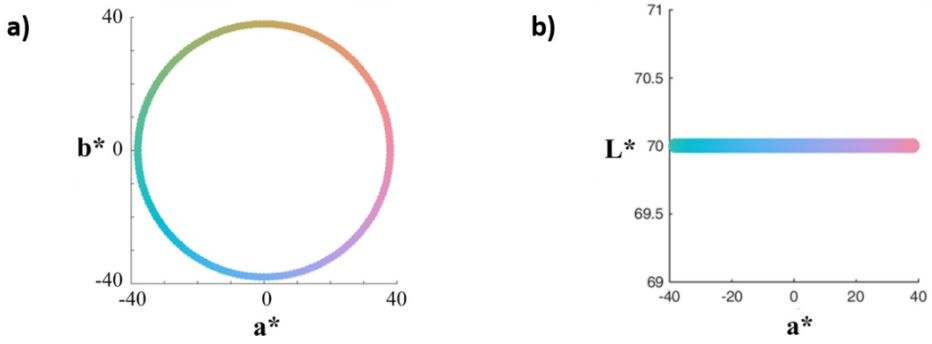

Figure 4-1 Color parameters in CIELAB space. **a)** Hue circle in a* and b* coordinates (a* center = 0, b* center = 0, radius = 38). **b)** Constant lightness (L*=70) for all color stimuli.

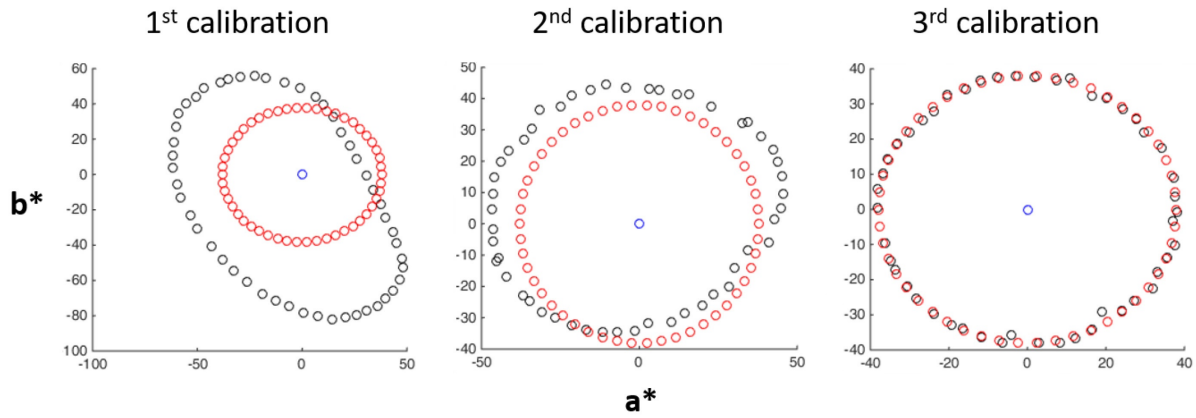

Figure 4-2 Color Calibration. With a spectroradiometer, parameters of 50 color stimuli were meaured and altered in iterations, to approximate the theoretically chosen CIE L*a*b* values on MRI monitor (monitor size: 1600 pixels * 1200 pixels). Blue dots: a* & b* center; black dots: desired hue values in CIELAB space; red dots: measured hue values in CIELAB space using the spectroradiometer.

An MRI-compatible spectroradiometer (JETI spectraval 1501) was employed to measure L*a*b* values of each of the 50 generated colors, and to calibrate these parameters on different screens (the MRI monitor for the MRI session and the computer screen for the behavioral

session). More specifically, we first calibrated for the reference gray color (white point), which approximates XYZ ratio of 1:1:1. Then using this gray as background color, each color stimulus was measured and changed in multiple iterations to minimize the discrepancy to theoretically chosen L*a*b* values (**Figure 4-2**).

4.2.3. Experimental Design

Three 2-h fMRI sessions for the delayed estimation (DE) task and one 2-h fMRI session for the undelayed estimation (UDE) task were completed by each subject. Afterwards, participants performed one approximately 90-min behavioral session which includes a category naming (CN) and a category identification (CI) task. Five sessions were conducted on different days, but within the same month, strictly following this sequential order: the DE task, the UDE task, and finally the CN and CI tasks. After the last fMRI session, participants also completed a 2-page questionnaire regarding their strategies for completing the working memory task. All experimental tasks were coded using PsychToolbox-3 (http://psychtoolbox.org/) and Matlab 2014b (Mathworks, Natick, MA).

Delayed Estimation Task

In the delayed estimation (DE) task, subjects memorized the target hue during the delay period and then reported on a hue circle. A trial (**Figure 4-3**) started with the sequential presentation of two color samples in the middle of the screen, followed by a retro-cue (either '1' or '2') at the center of a grey circle. The sample stimuli were concentric sinusoidal gratings within a circular aperture changing from the central gray point to the sample color, which drifted at a constant speed in a random direction: either inward or outward (Brouwer & Heeger 2009). The retro-cue informed subjects which of the two sample stimuli should be memorized for the rest of the trial, for example cue ‘1’ indicating the first presented color sample being relevant (cued sample) and the second sample being irrelevant (uncued sample) in this trial. The retro-cue was followed by the presentation of a blank screen (with only the fixation point) for 9.5 seconds, resulting in an overall delay of 10 s for memorization of the cued stimulus. Then a color wheel composed of all 50 color samples was presented in the center of the screen. Subjects were asked to mark the color wheel to best match the memorized sample within 4 s, by scrolling the trackball from

the screen center (cursor as a white dot) onto the color wheel (cursor as a white rectangular box), and by clicking the left button to confirm the choice. Once the selection was confirmed, both the color wheel and the response remained on the screen until the end of 4s. The color wheel was rotated by random degrees in each trial, thus to avoid motor preparation correlating with the hue position. Subjects were required to fixate throughout the trial.

The duration of one trial was either 18 s or 20 s (on average 19 s), including an inter-trial interval (ITI) of 2 or 4 s (on average ITI = 3 s). A run was comprised of 50 trials in random order, with each of the 50 sample stimuli presented once as the cued stimulus. Three fMRI scanning sessions resulted in altogether 16 runs and 800 trials for the delayed estimation task per subject. Before the first scanning session, subjects were trained for half an hour with feedback on their responses.

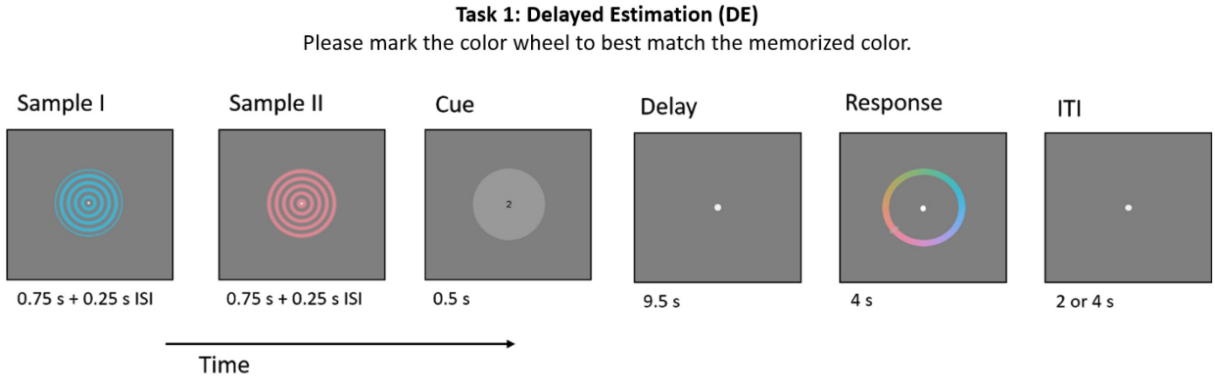

Figure 4-3 Experimental task 1: the delayed estimation task (working memory task; three fMRI sessions).

Undelayed Estimation Task

In the undelayed estimation (UDE) task, the delay period was removed, and subjects reported on the circular color wheel to best match the presented color sample. This task is comparable to the delayed estimation task except that the color wheel is presented at the same time as the sample stimulus, which can provide a direct contrast between working memory and perception processes (Gold et al., 2010; Brady et al., 2013; Bae et al., 2014, 2015; Souza et al., 2014). A trial (**Figure 4-4**) began with the presentation of a sample stimulus in the middle of the screen. Sample stimuli were concentric sinusoidal gratings that drifted at a constant speed in a random

direction: either inward or outward (same as in the delayed estimation task). Shortly afterwards, a color wheel faded in, while the sample remained shown (the color wheel started to appear 500 ms later than the sample presentation; the fade-in completed within 350 ms). The fade-in operation was to minimize the distracting effect of the color wheel from interfering with the subjects' focus on the stimulus. The color wheel consisted of all 50 sample colors, randomly rotated in every trial to avoid color-position association. Subjects were asked to click on the color wheel to select the color most similar to the sample color. The sample stimulus, the color wheel, and the response remained on the screen until the end of the trial (4 s long).

The next trial started after an inter-trial interval of 2 or 4 s (on average ITI = 3 s). A trial was thus either 6 s or 8 s (on average 7 s), and a run consisted of 50 or 100 trials. Altogether 700 to 800 trials were conducted for the undelayed estimation task per subject. Throughout the experiment, subjects were required to fixate at the fixation point.

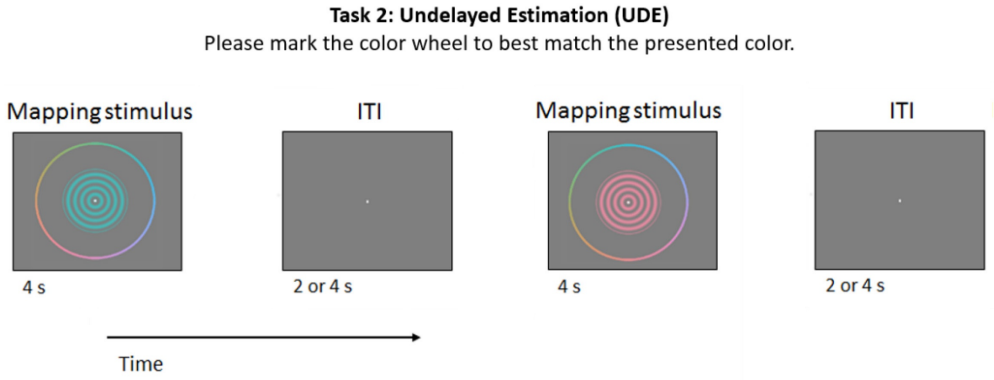

Figure 4-4 Experimental task 2: the undelayed estimation task (perception task; one fMRI session).

Category Naming and Identification Tasks

A pair of behavioral categorical tasks, category naming (CN) and category identification (CI) tasks, were performed in order to evaluate the categorization preferences of subjects (Witzel and Gegenfurtner, 2013; Bae et al., 2015). The tasks were conducted in a dark behavioral lab with either a keyboard or a mouse, after the completion of all fMRI sessions (DE and UDE tasks). Subjects had no time pressure as the next trial only started after they completed the response of this trial.

In the category naming task (**Figure 4-5 left**), a list of seven common color names including 'blue', 'pink', 'green', 'purple', 'orange', 'yellow' and 'red' was shown next to the sample color. These chromatic color terms were selected based on Berlin and Kay's eight basic color categories (Berlin and Kay, 1969), with 'brown' excluded, because both Bae's study (Bae et al., 2015) and our pilot study showed a very low usage frequency of the term to describe study colors in the color naming task. Subjects were asked to select the term that best described the color stimulus by pressing the up or down button on the keyboard, and to confirm their choice by pressing enter. The order of terms as well as the initial position of the cursor were randomized in each trial to minimize position bias. The sample size (trial number) is chosen based on previous work (Bae et al., 2015) but increased considerably: six subjects completed category naming evaluation 12 times for each of the 50 color stimuli, while four subjects evaluated each stimulus 9 or 6 times (due to time constraints of the behavioral session).

In the category identification task (**Figure 4-5 right**), subjects were required to mark the color wheel to identify the best example of each of the seven terms from the color term list. By pressing the left button of the mouse, they could confirm the color selection. The color wheel was rotated by random degrees in each trial to prevent association between the position and the color. A large sample size (trial number) was acquired in comparison to previous work (Bae et al., 2015): six subjects completed 90 category identification evaluation for each category term, while four subjects evaluated each term 60 times (due to time constraints of the behavioral session).

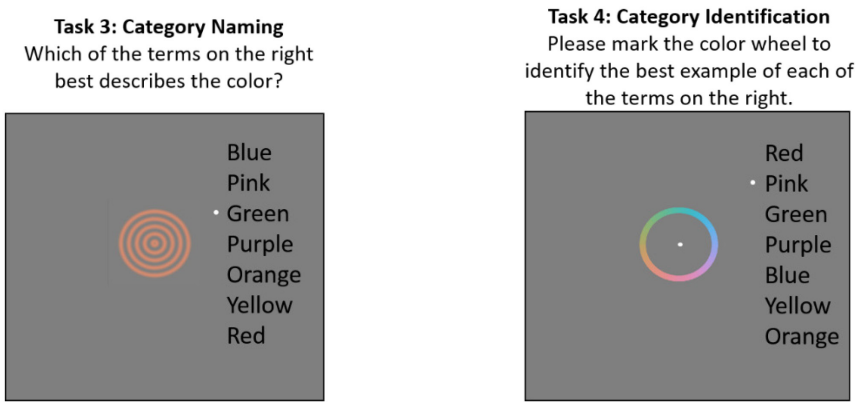

Figure 4-5 Experimental task 3: category naming and experimental task 4: category identification.

4.2.4. Data Acquisition

MRI data were acquired on a 12-channel Siemens 3 Tesla TIM-Trio scanner at the Berlin Center for Advanced Neuroimaging (BCAN; see section 2.1 for more about basics about fMRI). At the beginning of each scanning session, a high-resolution T1-weighted magnetization-prepared rapid gradient echo (MPRAGE) anatomical volume was collected (192 sagittal slices; repetition time TR = 1900 ms; echo time TE = 2.52 ms; flip angle = 9°; FOV = 256 mm). For acquisition of functional BOLD imaging, T2*-weighted echo planar images (EPI; 32 contiguous slices; TR = 2 s; TE = 30 ms; voxel size = 3x3x3 mm; matrix size = 64 × 64 × 32; slice gap = 0.6 mm; descending order; flip angle = 90°; FOV = 192 mm) were recorded covering the whole neocortex. Every trial was time-locked to the start of an EPI acquisition. For the delay estimation task, 478 EPI scans were collected per run, and altogether 7648 scans were acquired over 16 runs per subject. For the undelayed estimation task, 2450 to 2800 functional scans were recorded per subject.

Behavioral responses were collected via MRI compatible trackball in fMRI sessions, and via keyboard or mouse in a dark behavioral lab in behavioral sessions. Furthermore, subjects were asked to fill out a questionnaire after finishing all experimental tasks, which evaluates their strategies for memorizing color samples. Various strategies were listed in the questionnaire, including memorizing the stimulus by using description words, giving them name/code/number, or using acoustic, semantic and visual information, and association with a related emotion, action, touch, smell, or taste.

4.2.5. Anatomical Regions of Interest

In this study, fMRI data was analyzed within three regions of interest (ROIs) in visual areas V1, V4, and VO1 (see section 2.3.2.2 for more about ROI-based analysis). We focused the fMRI analysis on these three ROIs because they have been proven to exhibit selective neural response to color (Meadows, 1974; Zeki, 1974; McKeefry and Zeki, 1997; Bartels and Zeki, 2000; Brewer et al., 2005; Solomon and Lennie, 2007; Brouwer and Heeger, 2009; Riggall and Postle, 2012). These ROIs (**Figure 4-6**) were delineated based on high-resolution anatomical probabilistic maps estimated by Wang and colleagues (Wang et al., 2015), who superimposed individual maps of a large population of subjects (N=53) that were acquired through standard

fMRI paradigms of retinotopic mapping, and then transformed these maps into the standard MNI volume space (Collins et al., 1994; Wang et al., 2015). Every voxel within these probabilistic maps exhibits a group-average likelihood of belonging to a ROI, which varies between 0 and 100%. Notably, a large variation in probability exists across ROIs; for example, the maximal probability of a voxel belonging to early visual cortex (V1) and to intraparietal sulcus (IPS) is respectively 100% and 44% (Wang et al., 2015). These group-level probability maps can generally be applied to adult human brains.

The high-resolution probability maps were further processed to obtain the binary maps for every individual subject. First, these probability maps were deformed into the brain space of individual subjects by unified segmentation (Ashburner and Friston, 2005; by using 'inverse normalization' paradigm in SPM12). Then, the maps on the left and right hemispheres were collapsed. A map of the V1 region was obtained by combining ventral and dorsal maps of the early visual cortex (V1v and V1d). V4 and VO1 are adjacent brain regions and respectively refer to the fourth human visual field map hV4 (designated human V4 due to the unclear homology to macaque V4) and the ventral occipital cortex VO1. Additionally, a mutual exclusion rule was also applied so that every voxel in the brain only had probability value in the single ROI, out of all other ROIs, in which it had the highest probability. Furthermore, maps were thresholded to exclude voxels with a lower than 10% probability of belonging to a ROI, while other voxels were included. The thus acquired binary maps depict which voxels belong to ROIs of V1, V4 and VO1 for each individual subject.

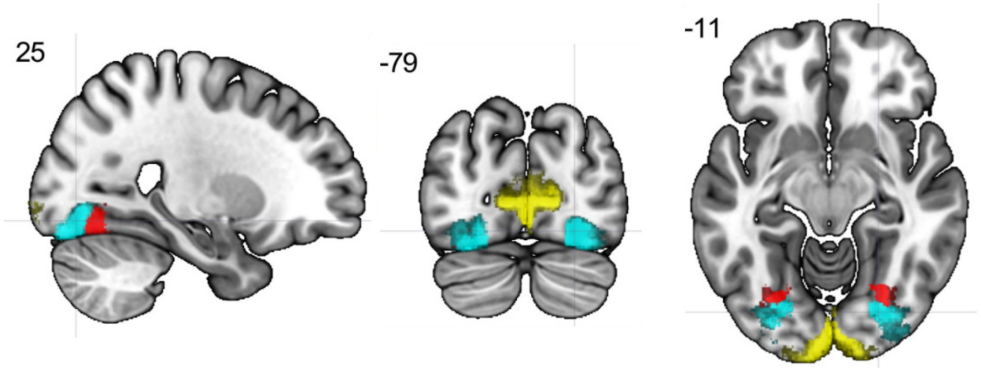

Figure 4-6 Three regions of interest (ROIs) in retinotopic probability maps, displayed in MNI volume space (at MNI coordinate of [25, -79, -11], based on Wang et al., 2015). Each color-coded region depicts a ROI across both hemispheres: V1 (including ventral and dorsal V1) is marked in yellow, V4 in cyan, and VO1 in red. This figure is visualized using MRIcroGL 64 (https://www.mccauslandcenter.sc.edu/mricrogl/home).

4.2.6. fMRI Preprocessing

The fMRI analysis was conducted by using the SPM12 software (Friston et al. 1994) and the cvManova toolbox (Allefeld et al. 2014). Additionally, we developed the code based on Sprague's IEM-tutorial examples (see https://github.com/tommysprague/IEM-tutorial; Ester et al., 2015).

The acquired images were first converted from DICOM format to a SPM compatible format of NIfTI. Next, all functional images belonging to one subject were realigned and resliced in order to correct head movement within and between runs. Then, the anatomical image was coregistered to the first functional image and further segmented (see section 2.2 for more about fMRI preprocessing).

Three scanning sessions for the delayed estimation task were concatenated (16 runs per session and 1 repeated measurement per run for each color sample). The estimated response from a single ROI can be depicted by a matrix of dimension m × n, with m referring to the number of voxels in the ROI and n depicting the number of repeated measurements of all sessions (n equaled number of runs multiplied by the number of repeated measurements for each color sample per run).

4.2.7. Color Encoding Basis Function

To estimate color-selectivity from spatially scattered and distinct response patterns of a population of voxels, an inverted encoding model (IEM) was employed (see section 2.3.2.4 for more about IEM). The basic assumption of this method was that various neurons within a voxel responded selectively to colors with different preferences, and the population response of neurons together covered approximately the whole color space (Brouwer & Heeger, 2009). Furthermore, it was assumed that a linear relationship existed between the response of a voxel and the sum response of all neurons in that voxel (Brouwer & Heeger, 2009). To characterize the selective neural response to color stimuli, encoding basis functions (BFs; or channels or filters) were constructed. There exists a one-to-one and invertible transformation from the color

stimulus to the channel output. The color selectivity of not only a neuron but also a voxel can be characterized as the weighted sum of a set of color-selective channels. By transforming feature space into channel space, it is possible to decode a large number of color samples.

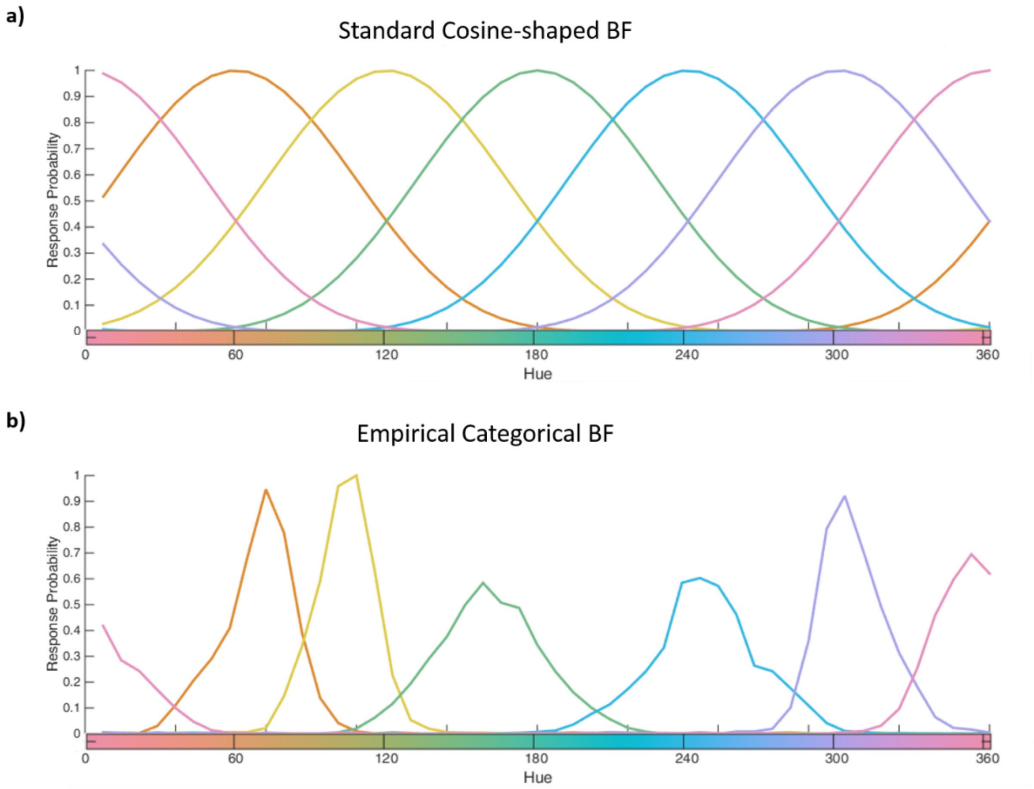

Figure 4-7 Two types of encoding models, each consisting of a set of six basis functions. The voxel response to color was modeled **a)** as classic half-wave rectified cosine functions to the power of six, which held low-level visual information and were evenly distributed over the hue space; **b)** as categorical mixed channels based on empirical data from a pair of categorization tasks, in order to characterize categorical neural representation.

Two types of basis functions were employed in this study to characterize the selective neural response to color information. The first type was a cosine-shaped 'steerable filter' (see section 2.3.2.4 for more about steerable filter), which was used by multiple previous IEM studies to model the selective neural response to either orientation or color (Freeman and Adelson, 1991; Brouwer and Heeger, 2009; Ester et al., 2013, 2015; Sprague and Serences, 2013). In this study we utilized a set of six half-wave rectified cosine functions raised to the power of six (**Figure 4-7a**). It was half-wave rectified to avoid a negative response (replaced by 0), which simulated the threshold effect of action potential (Brouwer and Heeger, 2009). And it was raised to a high

power to make the channels narrower and more selective. Furthermore, it showed a shape similar to the von Mises distribution which approximated the shape of single-unit tuning functions of cortical neurons (Brouwer and Heeger, 2009; Ester et al., 2013). The six basis functions were equally shaped and evenly distributed over the circular hue space and could simulate the color-selectivity in some brain regions where, overall, a comparable number of neurons respond to each separate color hue. Thus, this type of channel was used to characterize low-level visual representation of color in the brain.

In this study, we developed a novel type of basis function to model the additional categorical neural representation of color (**Figure 4-7b**). This type of basis function was constructed based on empirically acquired color categorization preference data (described in the following paragraph). A set of six categorical basis functions were estimated by combining response probability distributions of two tasks (**Figure 4-8 left**). The color naming task estimated the probability of category terms best describing color hues, and the color identification task evaluated the probability of color hues as best examples of each color categories. These two probability distributions were respectively smoothed and normalized, so that the probability sum of every category equaled one. By averaging these two response probability distributions we acquired mixed response probability distributions (**Figure 4-8 middle**). By further normalizing these by the highest relative response value among all six channels, we finally acquired categorical basis functions (**Figure 4-8 right**).

A set of categorical channels were utilized to characterize the group-level categorical effect in some cortical regions, where the overall neural response to some color hues prevails over the other hues. In contrast to the evenly spaced cosine-shaped channels which characterize the evenly distributed visual neural representation, the unevenly spaced categorical channels could characterize the unevenly distributed categorical neural representation of color. By comparing these two types of basis functions, it is possible to infer the dominant neural representation. For the main analyses and results of this chapter, we employed individual-based categorical channels relying solely on empirically acquired categorization preferences of individual subjects. Additionally, average-based categorical channels relying on average categorization preferences across ten subjects were constructed and compared with individual-based categorical channels (result comparison see **Figure 4-19** and section 4.3.6).

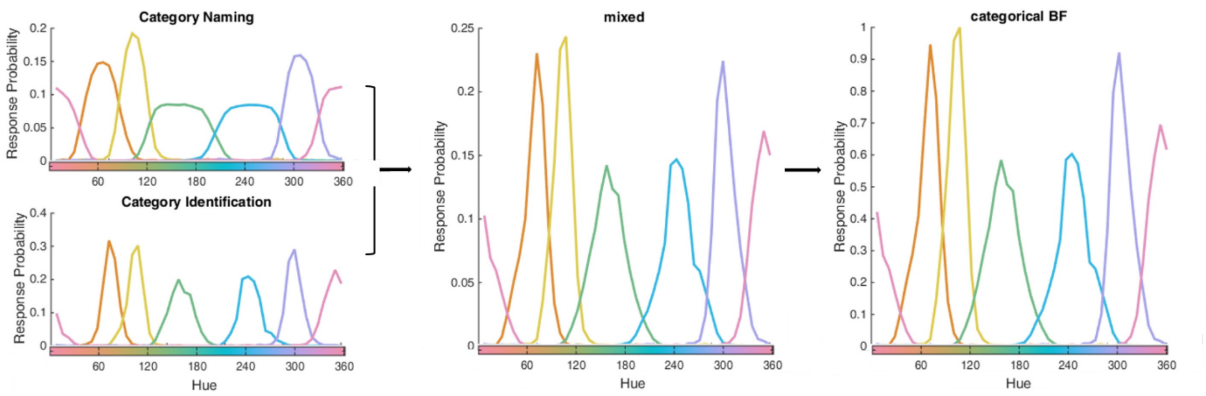

Figure 4-8 The categorical mixed probability distribution was estimated by combining the empirically acquired response probability distributions of the category naming task and the category identification task. The mixed distributions were normalized by the highest response value among all channels to estimate the categorical mixed basis functions. In this figure, we displayed average-based categorical channels relying on average categorization preferences across ten subjects, but individual-based categorical BF relying on the categorization preference of each individual subject was employed to estimate main results of this chapter (result comparison see section 4.3.6).

As described in section 2.3.2.4, the inverted encoding model is comprised of two stages. First, in the training stage, the voxel response is modeled by basis functions, whose weight is estimated in every voxel. Then, in the testing stage, the channel response is predicted based on brain data and previously estimated channel weights across voxels. In order to evaluate the strength of the sample representation in the brain, the distribution of the circular shift is estimated (Ester et al., 2013, 2015). However, the presumption of calculating the circular shift is von-Mises-shaped basis functions. Due to the irregular shape of the empirically acquired categorical basis function, an alternative approach to inverted encoding model is desired to evaluate the information measure held in the brain.

4.2.8. Multivariate Pattern Analysis

Here, we propose to combining the encoding basis functions (or channels) with a multivariate pattern analysis (MVPA), the cross-validated MANOVA (Allefeld and Haynes, 2014). The aim of cvMANOVA is to estimate information content held in the multivariate brain data that could differentiate between experimental conditions (see section 2.3.2.3 for more about cvMANOVA). To perform this combined analysis, sample hues were first projected into basis function space and then contrasted there for multivariate pattern analysis. The amount of

multivariate variance explained by the specific contrast between basis functions was estimated, which reflects the strength of color representation in the brain. The analysis was performed on a set of selected voxels within three regions of interest (ROIs): V1, V4, VO1 (see **Figure 4-6**). In each ROI, a scalar pattern distinctness D was estimated for every subject. This approach enables variable dimension reduction, which facilitates differentiation between a large number of experimental conditions with limited experimental data in MVPA. Furthermore, it alleviates the limitation of the inverted encoding model, so that the decoding performance of basis functions of irregular shape can be evaluated.

This combined approach starts with projecting sample hues into a set of six encoding basis functions (BFs) covering the whole circular color space. A generalized linear model (GLM) was constructed in preparation for the following cvMANOVA approach, where every sample color was modeled as a set of six parametric modulations representing six basis functions. For both delayed and undelayed estimation tasks, this procedure was performed and explained as below.

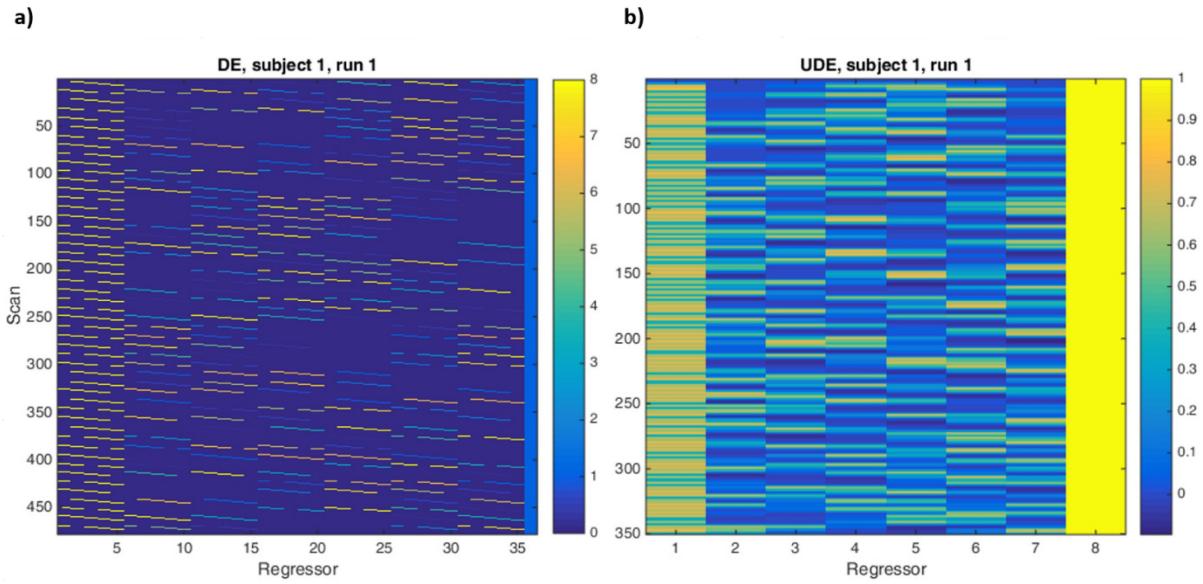

Figure 4-9 Design matrix for the first run of the first subject **a)** in the delayed estimation task, with 478 rows (478 scans) and 36 columns (35 regressors and 1 constant term); and **b)** in the undelayed estimation task, with 350 row (350 scans) and 8 columns (7 regressors and 1 constant term). In both tasks, the target hue was represented by six BF-based regressors and one stimulus-based average regressor. But while the 10 s delay period in the former task was modeled by five FIR bins, the 4 s presentation period in the latter task was modeled by one HRF bin, leading to altogether 35 regressors and 7 regressors in respective tasks.

In the delayed estimation task (**Figure 4-9a**), a GLM was built to estimate the memory-related brain activity in response to the cued color during the delay period. To represent the 10 s delay-period BOLD signal, five finite impulse response (FIR) regressors were employed in each trial (5 fMRI scans at a TR of 2 s). A design matrix was generated consisting of scans (rows) and regressors (columns). Every subject completed 16 runs with 50 trials and 478 scans per run, and thus the design matrix contained 478 rows per run. In every trial, the target hue was modeled by six BF-based regressors and one stimulus-based regressor. Every regressor was represented by five FIR bins, resulting in 35 regressors (plus one constant term, thus 36 columns) per run. The stimulus-based regressors (the first 5 columns representing the 5 FIRs with constant values) were estimated in SPM as the average of parametric values in the trial. Notably, no model interaction was performed because basis-function-related model parameters were not independent from each other.

The undelayed estimation task (**Figure 4-9b**) was modeled similarly, but the 4 s brain activity during the stimulus-presentation-period was represented by a canonical hemodynamic response function (HRF), which was time-locked to the stimulus presentation's onset. The design matrix of the first subject and the first run included 350 rows (350 scans). Similar to the delayed estimation task, each target hue was represented by six BF-based regressors and one stimulus-based regressor. However, each regressor was modeled by one HRF, leading to altogether 7 regressors (plus one constant regressor, thus 8 columns) in one run.

Next, these six encoding basis functions, instead of the 50 sample hues, were contrasted in a multivariate pattern analysis. A contrast matrix was introduced to specify the contrast. Except the stimulus-based average regressor, every pair of neighboring BF-based regressors defined in the design matrix was contrasted (BF 1 vs BF 2; BF 2 vs BF3; BF 3 vs BF 4...). In the delayed estimation task (**Figure 4-10a**), the transposed contrast matrix was comprised of 35 columns representing 35 regressors (six BF-based and one stimulus-based regressors, each in five FIRs bins) and 25 rows representing 25 contrasts (five contrasts between six BF-based regressors, each in five FIR bins). Similarly, the contrast matrix was constructed in the undelayed estimation task (**Figure 4-10b**). The transposed contrast matrix had 7 columns portraying 7 regressors (six BF-based and one stimulus-based regressors, each in one HRF bin) and 5 rows depicting 5 contrasts (five contrasts between six BF-based regressors, each in one HRF bin).

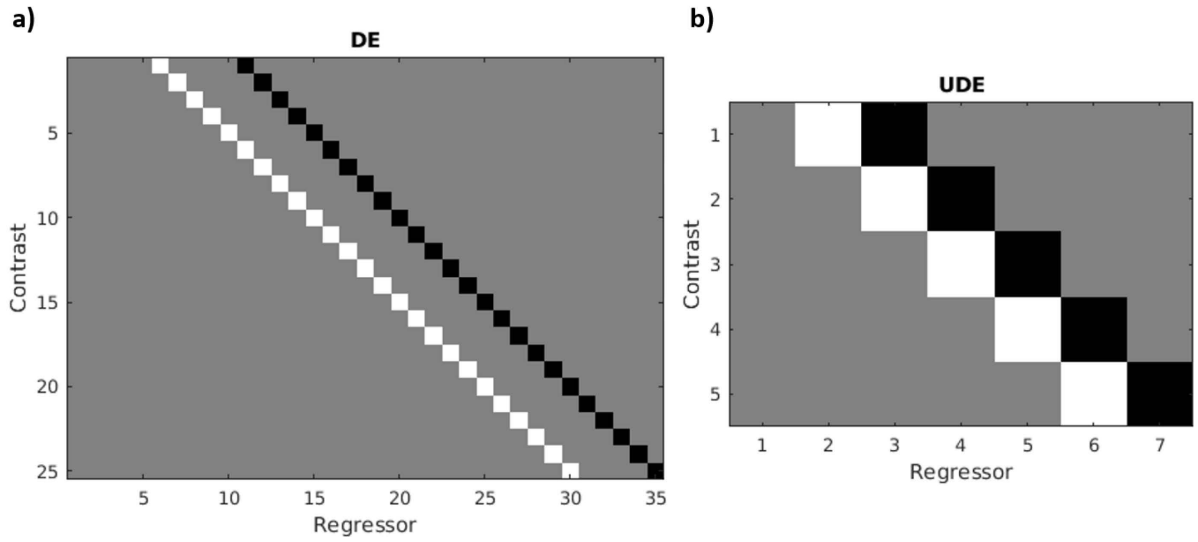

Figure 4-10 Transposed contrast matrix for the multivariate pattern analysis in the estimation tasks. Each row represents a contrast between two neighboring regressors (white, black, gray represent 1, -1 and 0 respectively). **a)** In the delayed estimation task, the transposed contrast matrix was comprised of 35 columns representing 35 regressors (6 BF-based plus one stimulus-based regressors, each in 5 FIRs bins). While the stimulus-based regressor was not for contrast, the six basis functions were contrasted in pairs (BF 1 vs BF 2; BF 2 vs BF3; BF 3 vs BF 4...) in each of the 5 FIR bins, resulting in 25 contrasts altogether. **b**) In the undelayed estimation task, the transposed contrast matrix had 7 columns representing 7 regressors (including 6 BF-based and 1 stimulus-based regressors, each in 1 HRF bin) and 5 rows representing 5 contrasts between every pair of neighboring BF-based regressors.

Furthermore, cvMANOVA estimated the amount of multivariate variance in the brain data specified by the contrast matrix that rejected the null hypothesis. The null hypothesis in the delayed estimation task was that there was no significant difference in any of the FIR bin in the multivariate pattern between six basis functions of the memorized color. The null hypothesis of the undelayed estimation task assumed that no significant change existed in the multivariate distribution between six basis functions of the perceived colors. The resulting pattern distinctness D reflects the information measure of color held in the brain.

To test the group effect across subjects, a nonparametric bootstrapping test was conducted. It estimates the group effect by random resampling (Efron, 1979; Bickel and Freedman, 1981; Singh, 1981). This method is not restricted by sample distribution shape (no problem with unknown or possibly non-Gaussian distributions) and can be applied to a small sample size, although a large sample size is preferred (Efron, 2003). In order to assess the statistical

significance , the bootstrap mean estimates of all resampling are compiled in a histogram and the bootstrap confidence interval (CI) is estimated (Davison and Hinkley, 1997). If 0 is not included in the (1-α) confidence interval, this suggests a significant effect at the threshold of α. For example, if α equals 5%, and if 0 is not included in the 95% CI, this suggests a statistical significance at the threshold of 5% (the p-value is regarded as less than or equal to 0.05). The random bootstrapping resampling process was repeated 100,000 times for 10 subjects in this study.

4.3 Results

4.3.1. Behavioral Results

Delayed and Undelayed Estimation Tasks

Behavioral performance of all trials were collapsed across sessions for each subject for both delayed and undelayed estimation (DE and UDE) tasks. We calculated the absolute hue distance between the response hue and the target hue – the *absolute error*, and examined the distribution of absolute errors in two estimation tasks. In the circular hue space, one hue distance equals 7.2 degree in a 360-degree circle, as in this study the CIE a*b* circular hue space was evenly divided into 50 hue samples. Compared to the UDE task (**Figure 4-11a,** in gray), in the DE task, all subjects exhibited broader distributions of absolute errors with lower peaks (**Figure 4-11a,** in black), suggesting an overall lower accuracy in the DE than in the UDE task.

Furthermore, the average absolute error was estimated for individual subjects as well as across ten subjects (**Figure 4-11b,** DE in black and UDE in gray**)**. In the DE task, the mean absolute error across subjects was 2.43 hue distance (equivalent to 17.53 degree in the 360-degree circular color space) with a SEM of 0.21 hue distance (1.53 degree). While in the UDE task, subjects responded on average with an absolute error of 1.36 hue distance (9.80 degree) and a low SEM of 0.05 hue distance (0.39 degree). It is clear that subjects responded more accurately in the UDE task than in the DE task, indicating a better performance in color perception than in color memorization. The behavioral data also show an interesting stimulus-specific bias pattern, which is investigated and discussed in 0.

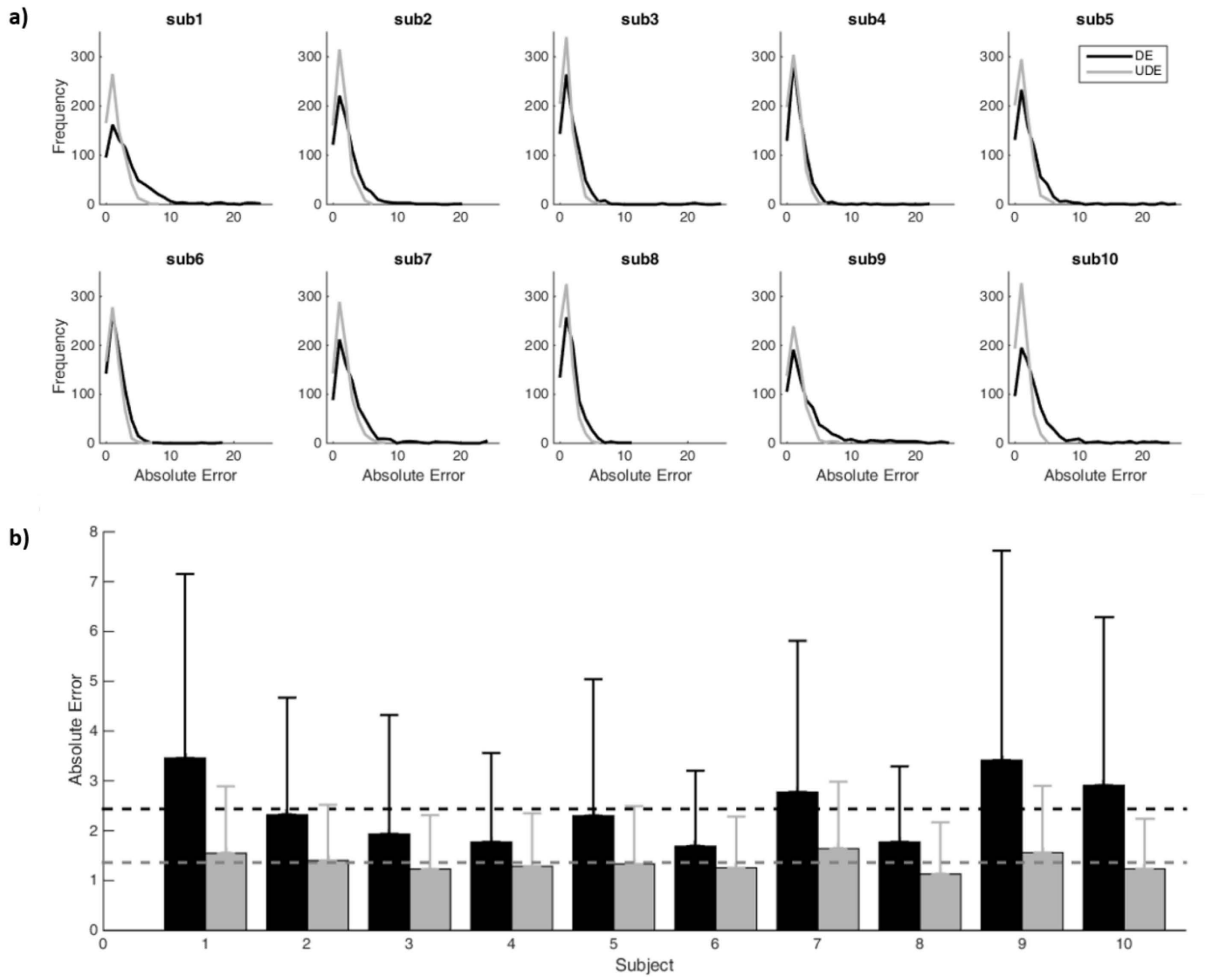

Figure 4-11 Behavioral performance of ten subjects in the delayed estimation (DE; in black) and the undelayed estimation task (UDE; in gray). **a)** The distribution of absolute errors (in hue) is illustrated for both tasks for each individual subject. **b)** The mean and the standard deviation of absolute errors (in hue distance) are estimated across all trials for each subject, which were respectively represented by bar graphs and error bars here. Additionally, the average absolute error across all subjects was illustrated as dashed lines for each task. In this chapter the CIE a*b* circular space was divided into 50 evenly spaced hues, where one hue distance equaled 7.2 degrees in a 360-degree circular color space.

Category Naming and Identification Tasks

In the category naming and identification tasks, we evaluated the color categorization preferences of subjects. The acquired behavioral responses were collapsed across all trials of all sessions for respective tasks.

In the category naming task, subjects marked the best category term to describe each hue (task see **Figure 4-5a**). Based on response frequency distributions (**Figure 4-12**), the boundary hues can be estimated. A hue that could be labeled with equal probability by adjacent category terms is considered a boundary hue. More specifically, we estimated by assessing the intersecting hues of response frequency distributions between neighboring categories. Furthermore, the category term 'red' was used in a notably low frequency to label a range of hues that mostly overlapped with the term 'pink' and the term 'orange' (**Figure 4-12**). Due to its redundant usage in naming hues, the term 'red' was excluded from further analyses in all tasks. The remaining six color category terms: 'pink', 'orange', 'yellow', 'green', 'blue', 'purple' were utilized to study categorical color representation, consistent with the ones used in previous works (Bae et al., 2015).

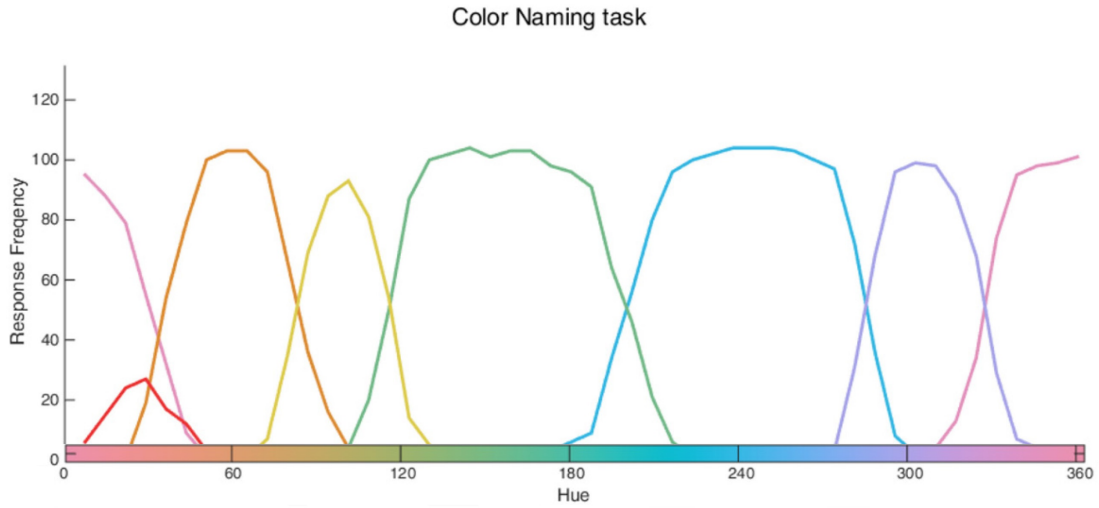

Figure 4-12 Collapsed results across all subjects of the category naming task, where subjects selected the best among seven category terms (including 'pink', 'red', 'orange', 'yellow', 'green', 'blue', 'purple') to describe each of the 50 sample hues. Please note that the term 'red' was excluded from further analyses in all tasks due to its redundant usage in naming hues.

The category identification task required subjects to mark the best hue exemplars of each color category (task see **Figure 4-5b**). Its response frequency distributions showed clear peaks and resembled the shape of the circular von Mises distribution (**Figure 4-13**). We estimated the focal hues from this task. A focal hue refers to a hue most frequently representing the best example of a color category. Here we assessed the focal hues by fitting the response frequency distributions with von Mises distributions. The estimated centers of these von Mises

distributions were considered focal hues. Because the ‘red’ category was excluded from further analyses due to its rare use in the color naming task, a certain discrepancy was left between the ‘pink’ and ‘orange’ categories.

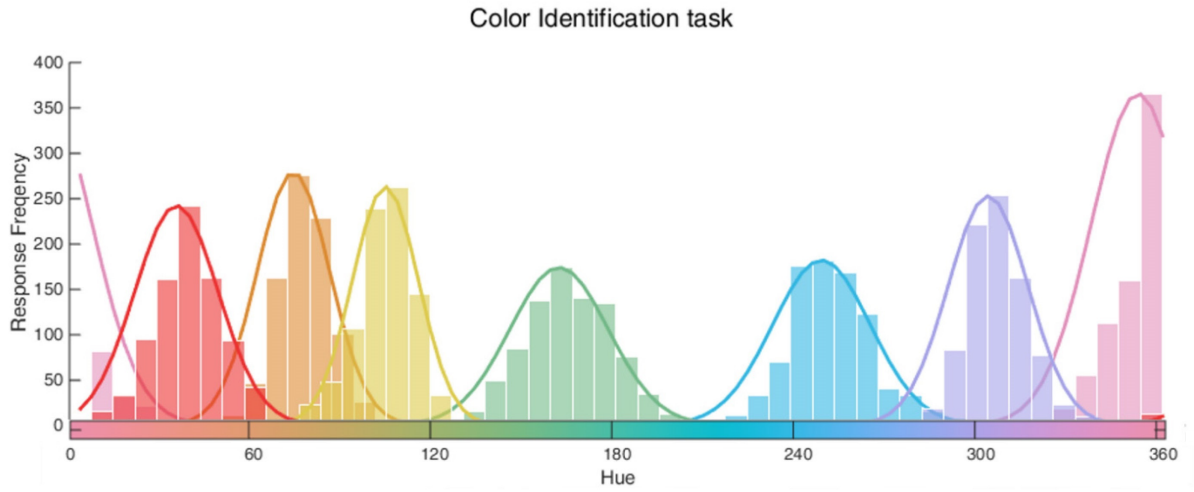

Figure 4-13 Collapsed results across all subjects of the category identification task, where subjects were asked to mark on the color wheel the best exemplar of each of the seven category terms (including ‘pink’, ‘red’, ‘orange’, ‘yellow’, ‘green’, ‘blue’, ‘purple’). We used von Mises distributions (curves) to fit the response frequency distributions (bars) of each color category. Please note that the term ‘red’ was excluded from further analyses in all tasks due to its redundant usage in the category naming task.

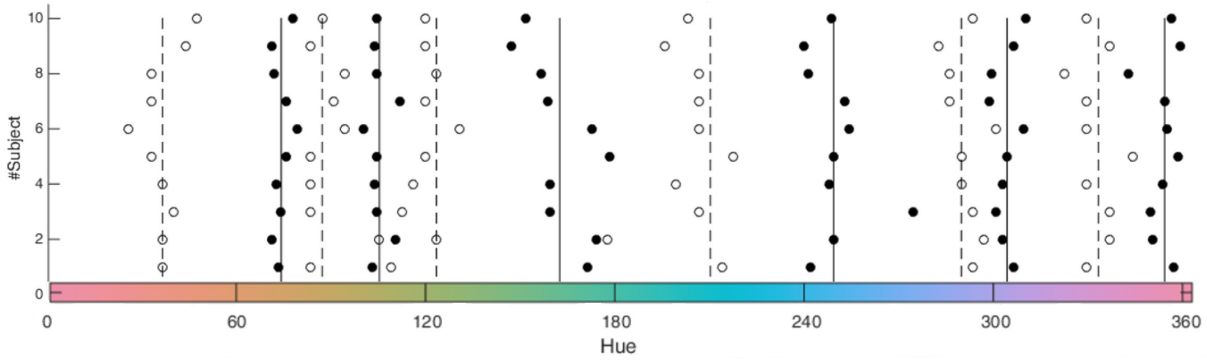

Figure 4-14 Six focal hues estimated based on the category identification task and six boundary hues calculated based on the category naming task (six categories: ‘pink’, ‘orange’, ‘yellow’, ‘green’, ‘blue’, ‘purple’, with ‘red’ excluded due to its notably low response frequency in the category naming task). While dots represent individual focal hues (filled dots) and boundary hues (empty dots) for each of the ten subjects; lines depict average focal hues (vertical black line) and boundary hues (vertical dashed line) across all subjects.

The response patterns of categorization pair tasks found in this study show consistency with Bae and colleagues' study (Bae et al., 2015) and bear resemblance to that of earlier categorization studies (Boynton and Olson, 1990; Sturges and Whitfield, 1997). For each individual subject as well as across ten subjects, a set of six boundary hues and six focal hues were estimated for six typical color categories including 'pink', 'orange', 'yellow', 'green', 'blue' and 'purple' (**Figure 4-14**). As illustrated, the difference in focal and boundary hues can be observed between individual subjects, particularly in terms of 'green' and 'blue' color categories, we see substantial individual difference. Based on both individual-level and group-level data of categorization pair tasks, categorical basis functions were constructed to characterize selective neural response to color (result comparison see **Figure 4-19**).

4.3.2. Questionnaire Results

After finishing all fMRI sessions, every subject was asked to fill out a questionnaire about the delayed estimation task. It contained statements describing strategies for the short-term memorization of the target color during the delay period. Subjects were required to rate how accurately each statement described their strategy for the memory task (0: applies not at all; 7: applies fully throughout the task). The order of statements was randomized for every subject.

Fur illustration purposes, the rating frequency distributions for each of the 12 statements are displayed in the order of descending median rating scores (**Figure 4-15**). Subjects memorized target color samples most frequently by employing verbal strategies such as 'by using words to describe them' or 'by giving them some name, code or number'. Another frequently used strategy was to memorize target colors visually 'as how they looked'. Subjects also reported memorizing color by their intensity or through an associated temperature. Other strategies such as memorizing as semantic, acoustic information, or through a related emotion, action, smell/taste, and touch were rarely utilized for color memorization.

More details regarding the verbal strategy were further asked in the questionnaire. While all six basic color categories we employed ("pink, orange, yellow, green, blue, purple", based on Berlin and Kay, 1969; Bae et al., 2015) were reported for usage, subjects made use of some additional basic color terms and several describing words. Most subjects reported using 7-16 basic color terms (including 'blue, green, purple, yellow, orange, pink, red, brown, turquoise,

mud, sea, watermelon…') combined with descriptive words including color intensity (like 'dark, deep, intense, bright, light'), emotion (like 'nice, pleasant, dirty, disgusting'), temperature ('cold, warm'), etc. Some participants also reported combining multiple basic words (such as 'green-yellow').

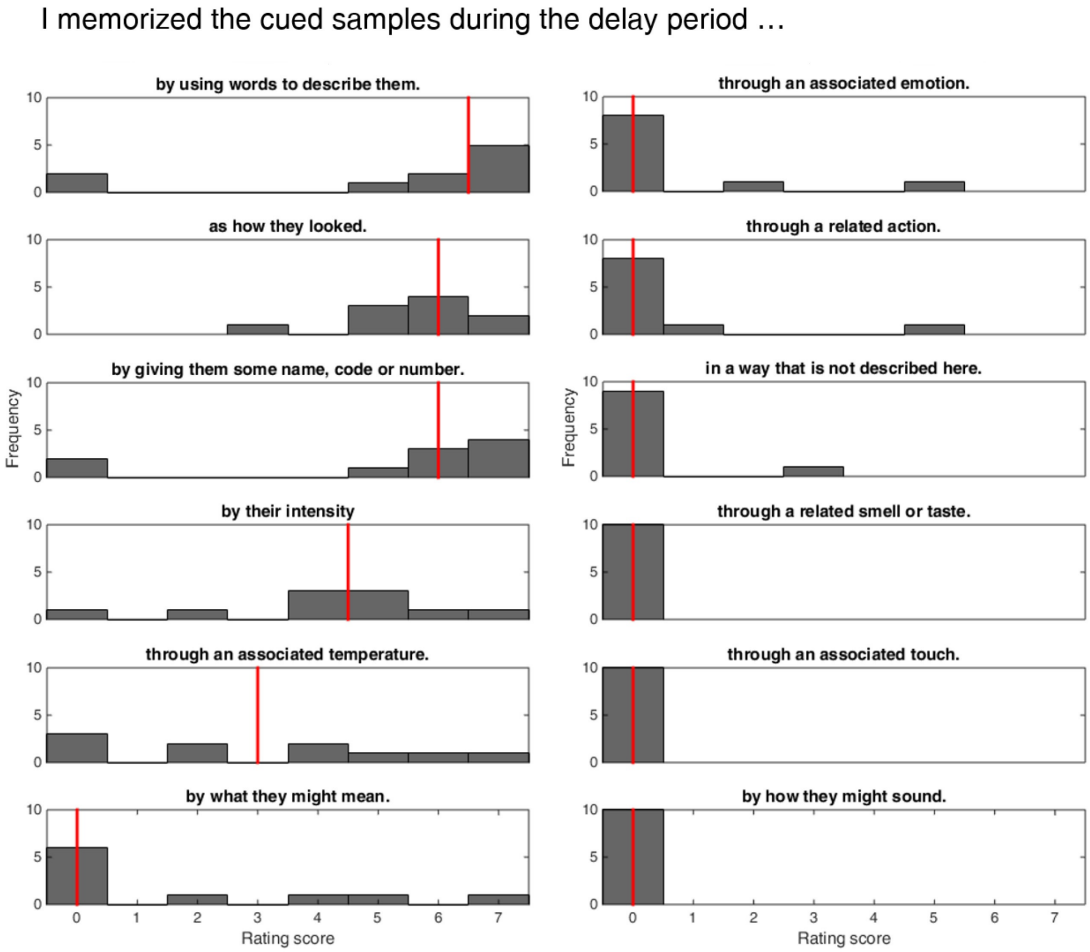

Figure 4-15 Frequency distribution of questionnaire results. After all sessions of fMRI experiments, every subject was asked to evaluate how accurately each statement described their strategy for the working memory task (0: applies not at all; 7: applies fully throughout the task). Statements were presented in random sequence to subjects, and ordered here in descending median rating. Red line: median rating of ten participants.

4.3.3. Color-specific Information Decoding

A multi-voxel pattern analysis, cvMANOVA, was combined with two types of encoding models (cosine-shaped and categorical basis functions) to estimate color-specific information from three regions of interest (ROIs): V1, V4 and VO1. Furthermore, to infer the group-effect, the bootstrap percentile confidence interval (CI) was estimated. If 0 is not included in the 95% CI, this indicates a significant effect (the p-value is regarded as less than or equal to 0.05).

To investigate the mnemonic representation of color in the brain, stimulus-specific content was estimated during the 10 s delay period (5 FIRs for 10 s time window) in the delayed estimation task. By using the classic cosine-shaped encoding model (half-wave rectified cosine basis function raised to the power of six), we found V1 and VO1 but not V4 exhibiting significant explained multivariate variance between target samples during the delay period across subjects (**Figure 4-16a;** bootstrapping with 100,000 random sampling of 10 subjects, multi-comparison corrected, 95 percentile confidence interval CI^{95} = [0.0013, 0.0293] in V1, CI^{95} = [-0.0038, 0.018] in V4, CI^{95} = [0.0002, 0.0082] in VO1). In contrast, when the categorical encoding model **(**categorical basis function based on empirical data from categorization tasks**)** was employed, all three ROIs were found to show significant information about the target color (**Figure 4-16a;** bootstrapping with 100,000 random sampling of 10 subjects, multi-comparison corrected, CI^{95} = [0.0011, 0.0486] in V1, CI^{95} = [0.0036, 0.0275] in V4, CI^{95} = [0.0048, 0.0134] in VO1).

To explore the perceptual representation of color in the brain, we conducted similar multivariate pattern analysis using categorical and non-categorical basis functions in the undelayed estimation task. Our decoding analysis focused on the 4 s stimulus presentation period (HRF with 4 s duration). All three ROIs showed significant multivariate variance between target colors when the non-categorical cosine-shaped encoding model was used (**Figure 4-16b;** bootstrapping with 100,000 random sampling of 10 subjects, multi-comparison corrected, 95 percentile confidence interval CI^{95} = [0.0198, 0.0724] in V1, CI^{95} = [0.0086, 0.078] in V4, CI^{95} = [0.0012, 0.0338] in VO1). In contrast, significant information for memorized color was decoded in V1 and V4 but not VO1 when the categorical encoding model was employed (**Figure 4-16b;** bootstrapping with 100,000 random sampling of 10 subjects, multi-comparison corrected, CI^{95} = [0.0185, 0.0783] in V1, CI^{95} = [0.0027, 0.0358] in V4, CI^{95} = [-0.0032, 0.0185] in VO1).

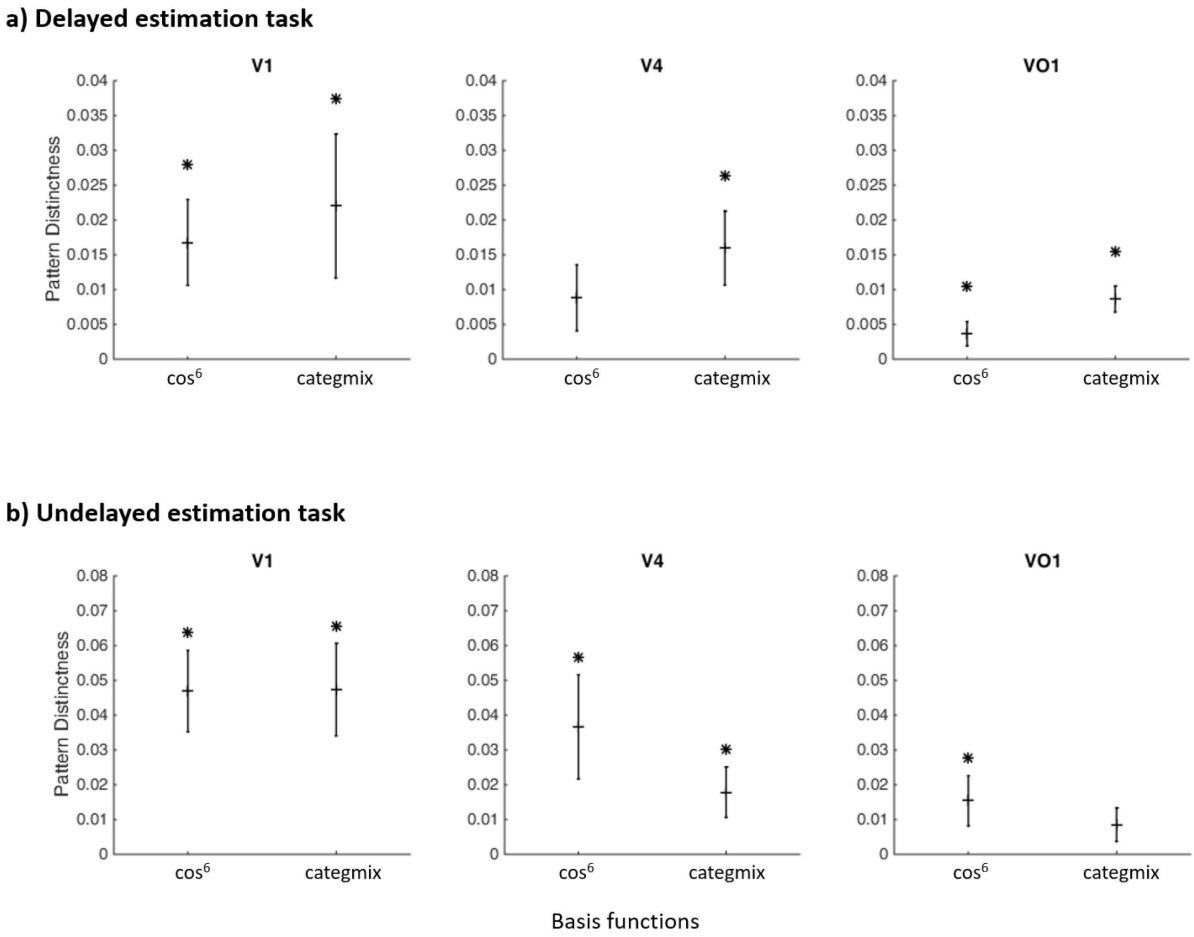

Figure 4-16 Decoding color-specific information in **a)** the delayed estimation task and **b)** the undelayed estimation task, using cvMANOVA and two types of encoding basis functions. Error bars refer to between-subjects SEM of the pattern distinctness D; * indicates significant effect, with 0 not included in the 95% bootstrap CI, 100,000 random sampling of the 10 subjects, $P < 0.05$, multi-comparison corrected.

4.3.4. Model Comparison

Next, the decoding results using two types of encoding models were compared, in order to assess the predominant information nature held in a brain area. To make the model comparison possible, we employed encoding models with an identical number of basis functions and normalized them with identical maximal amplitudes. While the cosine-shaped sensory encoding model characterizes sensory neural response, the categorical encoding model characterizes categorical neural response to color. We tested the statistical difference by contrasting these two models in a bootstrap test across ten subjects.

In the delayed estimation task, significantly higher information measure (pattern distinctness D) was found using the categorical model than using the visual model in V4 and VO1 (**Figure 4-17a**; with 0 not included in the 95% bootstrap confidence interval, 100,000 random sampling of the 10 subjects, $P < 0.05$, multi-comparison corrected), and no significant model difference was found in V1. While in the undelayed estimation task, no significant difference was found in any of our regions of interest (**Figure 4-17b**; in all ROIs: 95% bootstrap confidence interval included 0, 100,000 random sampling of the 10 subjects, $P > 0.05$, multi-comparison corrected). In an area specializing in carrying categorical information, significantly higher information would be decoded using the categorical than the sensory encoding model. From our results, one could infer that V4 and VO1 encode color information predominantly as categorical representation in working memory.

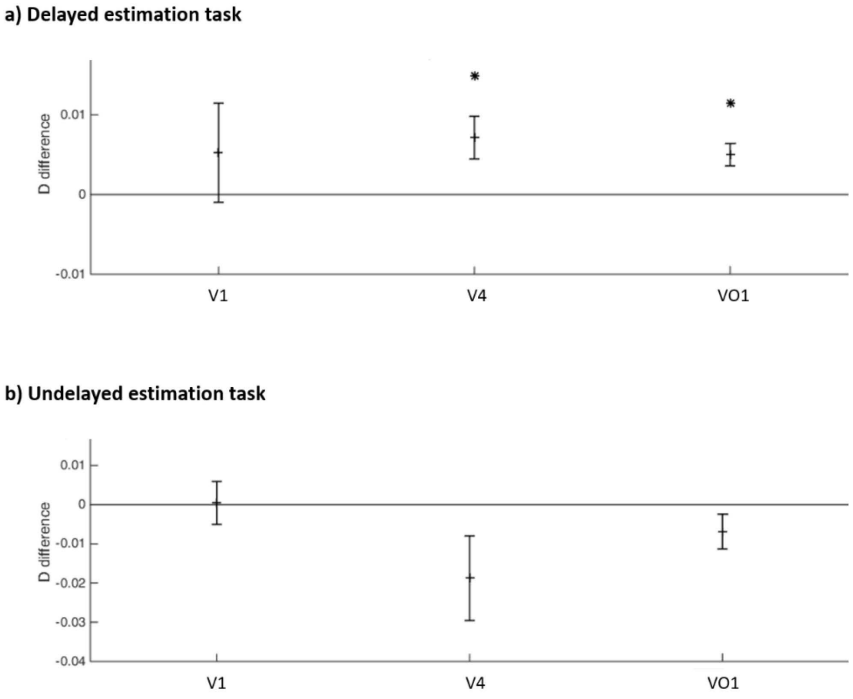

Figure 4-17 Model comparison in **a)** the delayed estimation task and **b)** the undelayed estimation task. In three ROIs, V1, V4 and VO1, we compared the pattern distinctness D of the sensory and the categorical encoding model, with positive D difference indicating a larger D using the latter compared to the former. Error bars refer to between-subjects SEM of the pattern distinctness difference; * indicates significant effect, with 0 not included in the 95% bootstrap CI, 100,000 random sampling of ten subjects, multi-comparison corrected.

4.3.5. Interaction Effect between Models and Tasks

In order to investigate how decoded information measures differ using conventional sensory versus empirical categorical encoding models between the delayed and the undelayed estimation tasks, we tested the interaction effect between models and tasks. However, it is not possible to directly compare information measure (pattern distinctness here) between tasks. On the one hand, this is due to the different level of mnemonic and perceptual representation strength of color (perceptual brain signals are often much stronger). On the other hand, this is because distinct response functions were utilized to estimate memory-related (5 FIRs for 10 s delay period) and perception-related (a HRF for 4 s stimulus presentation period) brain signals in respective tasks. Therefore, the pattern distinctness D was first standardized in every bootstrap sampling by averaging across sampled subjects and models before task comparison.

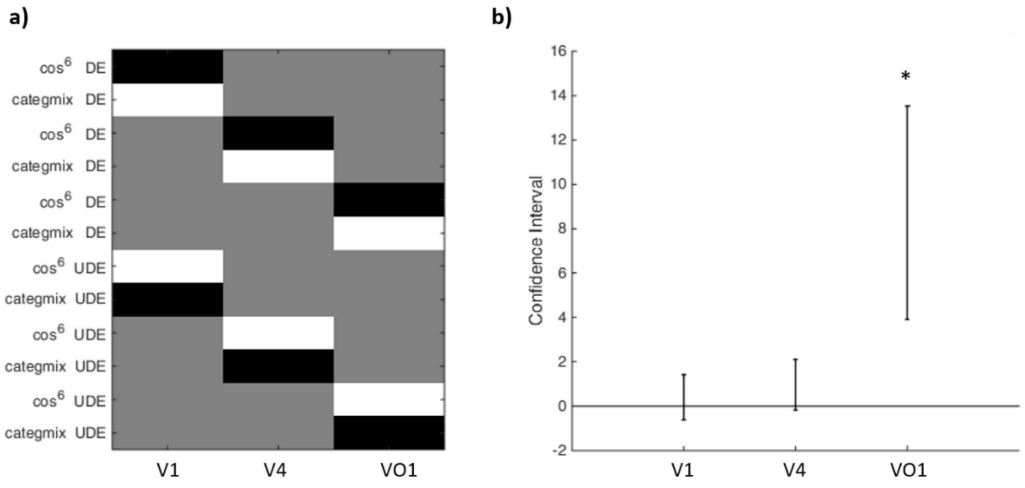

Figure 4-18 Interaction effect between encoding models and tasks in each of three ROIs. **a)** Illustration of three contrasts (white indicates 1, black represents -1, and gray refers to 0) targeting the interaction effect between encoding models (cosine-shaped versus categorical basis function) and tasks (delayed versus undelayed estimation task). **b)** The result of the interaction effect test in three ROIs. Bars indicate the 95% bootstrap confidence interval of the difference of the standardized pattern distinctness D, with 10,000,000 random resampling; * suggests significant interaction effect, with 0 not included in the 95% bootstrap CI, $P < 0.05$, multi-comparison corrected.

Next, the standardized D was compared in pairs (**Figure 4-18a**) between two encoding models (using cosine-shaped versus categorical basis functions) and two tasks (delayed versus undelayed estimation task) in three regions of interest. With 10,000,000 random bootstrap

sampling of ten subjects (**Figure 4-18b**), VO1 showed a significant interaction effect between models and tasks (multi-comparison corrected, with 0 not included in the 95% bootstrap confidence interval of the difference of standardized pattern distinctness D, $P < 0.05$), while V1 and V4 exhibited no significant effect (multi-comparison corrected, with 0 included in the 95% bootstrap confidence interval, $P > 0.05$).

4.3.6. Comparing Individual-based and Average-based Categorical Models

So far we have presented the main MVPA results of using a categorical encoding model, based on color categorization preferences of individual subjects (section 4.3.3 to 4.3.5). In this section, we further examine the MVPA results of using categorical encoding model based on average categorization preferences across ten subjects and compared these with findings using an individual-based categorical encoding model.

First, the average color categorization preferences were estimated by averaging individual preferences acquired in category naming and identification tasks. Based on these average preferences, the average-based categorical encoding model was constructed, which was repeatedly utilized to estimate color-specific information (pattern distinctness D) for every subject. Then, 100,000 bootstrapping resampling of 10 subjects were conducted (multi-comparison corrected) in order to estimate the group-level effect using the average-based categorical encoding model. In both delayed and undelayed estimation tasks, three ROIs exhibited similar information measures using average-based and individual-based encoding models (**Figure 4-19 a, b**).

Next, the difference between utilizing average-based and individual-based encoding models was statistically tested, respectively in the delayed and undelayed estimation tasks (**Figure 4-19 c, d**). With 100,000 random bootstrap sampling of 10 subjects, we found no significant model difference in either task (with 0 included in the 95% bootstrap CI, $P > 0.05$, multi-comparison corrected).

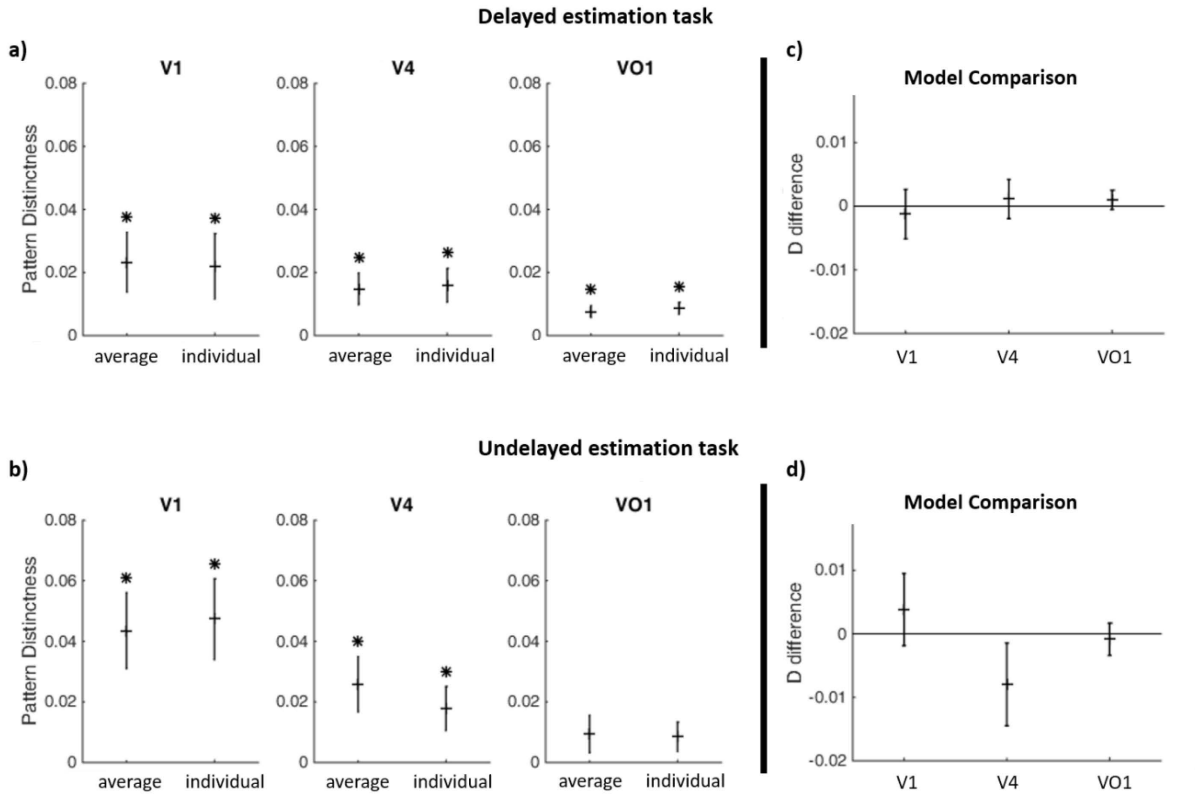

Figure 4-19 Comparison between average-based and individual-based categorical encoding models, respectively in the delayed (a and c) and undelayed (b and d) estimation tasks. **a), b)** illustrates the multivariate pattern distinctness D decoded from the brain using respectively average-based and individual-based categorical encoding models; while **c), d)** display the distinctness difference between using average-based and individual-based categorical models (positive D difference meaning higher D in individual-based model). Error bars refer to between-subjects SEM; * indicates significant effect, with 0 not included in the 95% bootstrap CI, 100,000 random sampling of the 10 subjects, multi-comparison corrected.

4.4 Discussion

In this chapter, I introduced a study that was comprised of delayed and undelayed estimation (DE and UDE) tasks conducted in the fMRI scanner (**Figure 4-3** and **Figure 4-4**), as well as category naming and identification tasks performed in the behavioral lab (**Figure 4-5**). Working memory and perception processes were assessed in the delayed and undelayed estimation tasks, respectively.

The key hypothesis to be tested in this study is that the neural representation of color working memory is dual-content. It is supported by our questionnaire result where subjects most

frequently reported memorizing target color by using verbal strategies, followed by a visual strategy (**Figure 4-15**). The hypothesis was further tested based on a critical assumption that two types of neural representations can be characterized by two types of differently constructed color-selective encoding models. A conventional cosine-shaped encoding model (Brouwer and Heeger, 2009; Sprague and Serences, 2013; Ester et al., 2015) and a novel empirical categorical encoding model based on empirically acquired color categorization preferences were employed to respectively characterize low-level visual and abstract categorical representations of color in the brain (**Figure 4-7**). While some claimed that neither the channel shape nor the space between channels has a significant influence on decoding performance (Freeman and Adelson, 1991; Brouwer and Heeger, 2009), we argue that channel construction could influence the decoding of specific types of information held in the brain. Encoding models were further combined with a MVPA approach to examine the information measure of color representation in each of the three ROIs that exhibited neural color selectivity: V1, V4, VO1 (**Figure 4-6**; Meadows, 1974; Zeki, 1974; McKeefry and Zeki, 1997; Bartels and Zeki, 2000; Brewer et al., 2005; Solomon and Lennie, 2007; Brouwer and Heeger, 2009).

First of all, both conventional cosine-shaped and empirical categorical encoding models were utilized to examine the mnemonic neural representation of color in the DE task. Using the cosine-shaped encoding model we identified significant low-level visual representation of color in V1 and VO1, while the using categorical model, we found categorical representation in all three ROIs (**Figure 4-16a**). Additionally, these encoding models were used to investigate perceptual neural representation of color in the UDE task. Significant visual representation was estimated in all three ROIs, while categorical representation is identified in all ROIs but VO1 (**Figure 4-16b**). Our results for perceptual color representation using the sensory encoding model are consistent with findings from a previous MVPA study on color vision (Brouwer and Heeger, 2009). The insignificant information measure could possibly but not necessarily be explained by the limited amount of neuroimaging data with insufficient statistical power. To summarize, we found three color-related cortical regions exhibiting neural representation of color information in both working memory and perception. Furthermore, by using two types of encoding models, the dual-content neural representation of color information is decoded in these brain regions. However, some regions showed a possible preference for one strategy, while others did not.

To find the predominant type of color representation in a brain region, the novel categorical encoding model was compared to the traditional sensory encoding model (**Figure 4-17**). During working memory in the DE task, V4 and VO1 exhibited significantly more stimulus-specific information using the categorical than the sensory encoding model, while V1 showed no significant difference (**Figure 4-17a**). This suggests a predominant categorical mnemonic representation of color in more anterior regions of visual areas: V4 and VO1, and a more balanced and reliable dual-content neural representation in V1. This dissociation between V4-VO1 and V1 has also been found in a previous PCA study (Brouwer and Heeger, 2009) that argued for a conversion from V1 for low-level neural representation (with a distorted color space) to V4 and VO1 with a clear cortical color space (e.g. a purple-blue is encoded between purple and blue).

During color perception in the UDE task, none of the three ROIs showed significant difference between categorical and sensory encoding models (**Figure 4-17b**). One possible explanation is the less frequent usage of categorical strategy in the UDE task. Although according to Bae's study on response patterns, both working memory and perception processes (DE and UDE tasks) involve dual-content representation of color (Bae et al., 2015), the proportion of categorical strategies could vary vastly. Our behavioral results showed that subjects completed the UDE task with clearly higher precision than the DE task (**Figure 4-11**). In a simple task where one can simultaneously see and compare the target hue with the response color wheel (UDE task), one can rely less on abstract verbalization and categorization of the visual stimulus. The low usage of categorical strategy might lead to a reduced categorical representation in all related brain regions during color perception. Thus, the UDE task could possibly serve as a contrast baseline (predominantly low-level visual neural representation of color) for studying the dual-content neural representation of color working memory.

To investigate whether the predominant categorical representation of color compared to the sensory representation is statistically different between DE and UDE tasks in a cortical region, the interaction effect was tested. Via pairwise comparison of standardized pattern distinctness, VO1 showed a clear interaction effect and its predominant categorical representation was statistically dominating in DE task (**Figure 4-18**). The finding implies that VO1 might be specialized in the working memory storage of categorical representation of color.

Next, we compared the categorical encoding model based on the color categorization preferences of individual subjects with the model based on average categorical preferences

across the ten subjects. The individual-based categorical encoding model was used for main analyses in this study. By testing the statistical difference, we asked whether the average-based or the individual-based categorical encoding model better characterizes the neural color-selectivity of the ten subjects. No significant difference was found in any ROI in both tasks (**Figure 4-19**). While the individual-based model better captures the individual difference in color categorization preference, the group-average model benefits from the ten-fold sample size. As shown in **Figure 4-14**, for most color categories, the ten subjects exhibited similar focal and boundary hues, while in some categories like 'green' and 'blue', notable individual difference can be observed.

To summarize, as anatomically moving from posterior to anterior, V1, V4, and VO1 displayed an elevation in categorical representations of color working memory. This finding implied a gradient of abstraction in the memorized content along the rostral-caudal axis of the brain. Furthermore, our new approach to characterizing separate neural representation forms of a feature by constructing distinct types of encoding models might provide an important and novel possibility for the examination of neural mechanisms.

A key step in this study is the construction of the categorical encoding model, which is primarily based on category terms utilized for category naming and identification tasks. Initially, seven color terms ('pink', 'red', 'orange', 'yellow', 'green', 'blue', 'purple') from the basic color category list (Berlin and Kay, 1969) were utilized, but the term 'red' was excluded from fMRI analyses (DE and UDE) due to its rare usage in the color naming task. This might influence the preference estimation of other categories, especially the terms adjacent to 'red', such as 'orange' and 'pink'. Furthermore, subjects reported employing a number of diverse color terms beyond these six basic color terms to complete the memory task. This could possibly lead to a less precise evaluation of categorization preferences in the color category tasks. A more sophisticatedly designed pair of category tasks, for example, with individually reported color terms, might lead to a categorical encoding model that better characterizes the categorical representation of color in the brain of individual subjects.

One potential limitation of this study is the subject number. Ten subjects participated in the experiments, each completing four 2-h scanning sessions and one behavioral session. Instead of inviting more subjects for a single session (for example 40 participants, one scanning session per participant), we recruited fewer subjects for multiple sessions (10 participants, four scanning session per participant). With equal scanning time in total, recruiting fewer subjects

reduces individual difference in brain anatomy and brain functions (Cosgrove et al., 2007). However, while the scanning trial number is large, the subject number of 10 is not optimal for the 2nd level statistics. This issue is especially important when the interaction effect between two times two conditions is regarded. To compensate for this, 10,000,000 bootstrap resampling were performed to estimate the confidence interval properly.

Chapter 5 Study III: Assessing Dual-content Representation of Color Working Memory Based on Response Patterns and a Probabilistic Model

Brief Summary of this Empirical Study

This study further tests the dual-content representation of color working memory. Similar paradigms (a delayed estimation task, a pair of color category tasks) have been utilized as in Chapter 4, but here, the research question is approached from a different angle. The first goal of this chapter was to examine whether subjects' responses exhibited a systematic color-specific bias pattern in the delayed estimation task. We employed 180 equally illuminated color samples that composed a whole circular color space as stimuli. The response bias exhibited a pattern apparently deviating from the uniform distribution, and was related to its distance to the category focal and boundary hues. In summary, the behavioral results showed an approximately systematic categorization effect.

The second goal was to implement a dual-content probabilistic model that combined the continuous visual representation with the discrete categorical representation of color. We found a significant correlation between the simulated and empirically acquired response bias pattern. For comparison, a similar working memory model based solely on visual representation was implemented, which exhibited an insignificant correlation with the experimental data. In conclusion, this study provides supporting evidence for a joint utilization of low-level visual and categorical representations in color working memory.

5.1 Introduction

To retain color in working memory is a challenging cognitive task and the mechanism of this process has received extensive attention. In the majority of previous studies, color is modeled

as estimates of hues on a continuous scale (Zhang and Luck, 2008). Recently, it has been proposed that color working memory also relies on categorical representation using color category terms (Bae et al., 2015). Shared among people from different language backgrounds, color category terms such as 'blue', 'pink', 'green', 'purple', 'orange', 'yellow', 'red', 'brown' are frequently utilized as basic color categories (Berlin and Kay, 1969; Bae et al., 2015). To some degree, the contribution of category terms to color working memory is comparable with how a compass helps us remember a path in dense woods. Typically, as we observe and memorize the visual details of the surroundings, we also remember the direction term (such as 'north', 'south', 'east', and 'west') showed by the compass at the same time.

The contribution of categorical representation in addition to visual metric representation of color information could result in certain biased response patterns in the delayed estimation task (Bae et al., 2015). A comparable case is the observation that categorical information such as landmarks can serve spatial working memory and cause bias patterns in response. Previous experiments showed that when subjects try to reproduce a dot on paper from memory, the aggregate results produce an approximate two-dimensional Gaussian response distribution throughout the paper, but the introduction of a boundary circle causes the subject to exclude all potential responses outside the boundary circle (Huttenlocher et al., 2000; Crawford et al., 2006; Duffy et al., 2010; Bae et al., 2015). The delayed estimation task is commonly used to study color working memory and to infer its structure based on the response bias (Wilken and Ma, 2004; Zhang and Luck, 2008; Bays et al., 2009, 2011; Fougnie et al., 2010; Fougnie and Alvarez, 2011; van den Berg et al., 2012; Bae et al., 2015), and a pair of categorical tasks are frequently employed to study color categorization preference of subjects (Witzel and Gegenfurtner, 2013; Bae et al., 2015). The estimated preference (such as the typical color of a category or the typical color-term to describe a color) can be utilized to examine and explain the delayed response bias patterns.

While typical mathematical models of color working memory estimate color as a continuous visual metric representation (Zhang and Luck, 2008), a novel model was proposed that additionally included the categorical representation of color (Bae et al., 2015). Based on the CATMET model from Bae and colleagues, we developed and implemented a probabilistic dual-content model to simulate color working memory. It models color information not only as visual metric estimates using von Mises distributions (Huttenlocher et al., 2000), but also as color category terms through probabilistic assignment. The von Mises distribution is the circular

analogue to the normal distribution and is frequently used to simulate circular feature space (Mardia and Jupp, 2000). The categorical pathway performed a probabilistic color category assignment based on empirical results acquired in the categorization pair tasks.

In this behavioral study, ten subjects were recruited to complete a delayed estimation task and a pair of color categorization tasks. A set of 180 color samples with equal lightness were employed as stimuli which form a whole circular hue space with an equal spacing of two degrees between neighboring samples. While similar paradigms have been utilized by Chapter 4, this chapter focuses on 1) analyzing and discussing the behavioral patterns in color working memory; 2) estimating boundary and focal hues; 3) modeling the color working memory by employing a dual-content probabilistic model.

5.2 Methods

5.2.1. Participants

Ten subjects (7 female, 3 male; aged 18-40; no overlap with participants in study II) with normal or corrected to normal visual acuity and no color blindness took part in the experiments. This study was granted ethical approval by the Charité ethics committee and all subjects gave informed consent.

5.2.2. Stimuli

A set of 180 color samples were generated in Commission Internationale de l'Eclairage (CIE) LAB space. They had a constant lightness (L*=70) and were equally spaced in the A*B* space (a*center = 0, b*center = 0, radius = 38; **Figure 4-1**). The L*A*B* parameters were selected based on a previous study on color working memory (Bae et al., 2015). The CIELAB space is a nonlinear conversion of the CIEXYZ space which was designed to be 'device-independent' and perceptually more uniform (Commission Internationale de l'Eclairage, 1986). These 180 color samples form a complete color wheel (360 degrees) with an equal spacing of two degrees between neighboring hues. For categorization tasks, a set of eight commonly used color

category terms consisting of ‘blue’, ‘pink’, ‘green’, ‘purple’, ‘orange’, ‘yellow’, ‘red’ and ‘brown’ was employed for evaluation (Berlin and Kay, 1969).

5.2.3. Experimental Design

Participants were asked to complete three tasks in the following sequential order: (1) the delayed estimation task, (2) the category naming task and (3) the category identification task. All experimental tasks were coded using PsychToolbox-3 (http://psychtoolbox.org/) and Matlab 2014b (Mathworks, Natick, MA). The paradigms are similar to those used and described in detail in Chapter 4. The number of trials for these three tasks was chosen based on previous studies (Bae et al., 2014, 2015).

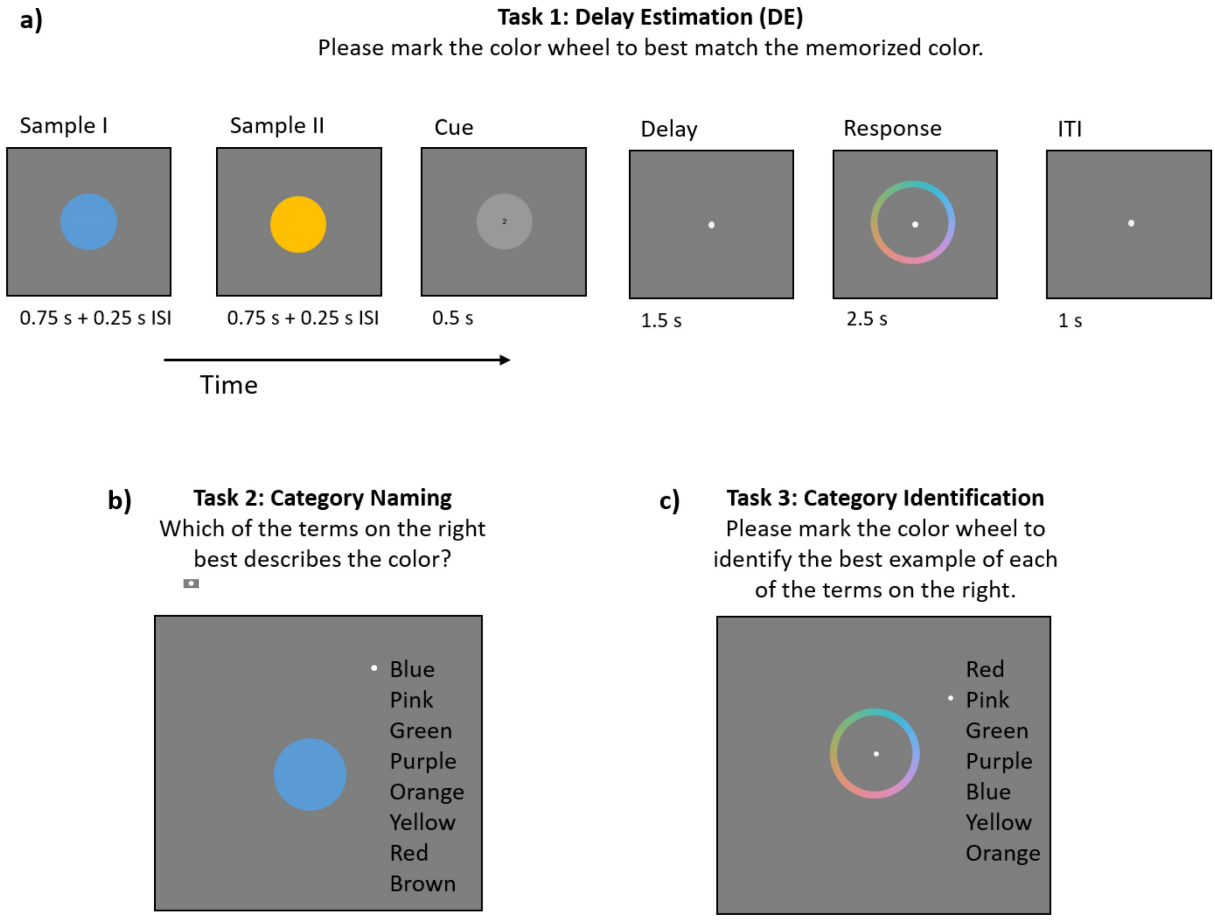

Figure 5-1 Experimental paradigms for **(a)** the delayed estimation task, **(b)** the category naming task and **(c)** the category identification task.

In the delayed estimation task, subjects were asked to memorize the target color sample and to mark the continuous color wheel to match the memorized color (see **Figure 5-1a**). The paradigm was similar to Chapter 4, except that 1) samples were presented as still color circles with no inward or outward drift; 2) 180 instead of 50 sample hues, which formed a complete circular color space, were employed as stimuli; 3) a shorter delay period (2 s) and a shorter response time (2.5 s) were used; 4) subjects sat in front of a computer screen and responded using a mouse instead of lying in the MRI scanner and responding with MRI-compatible 2*2 button boxes. Each of the ten subjects completed eight runs, with 90 trials per run. This led to 720 trials per subject, and 40 trials across participants for each of the 180 sample colors. Sample colors were presented in random order and the color wheel was randomly rotated.

In the category naming task, subjects were asked to select the term from the basic category list (‘blue’, ‘pink’, ‘green’, ‘purple’, ‘orange’, ‘yellow’, ‘red’, ‘brown’) that best described the represented color sample (see **Figure 5-1b**). The paradigm was similar to Chapter 4, except that 1) samples were shown as still color circles without inward or outward drift; 2) 180 sample hues were presented. Each subject completed two runs, with 360 trials per run, resulting in four trials per sample color per subject. Both sample colors and color terms were presented in random order in every trial.

Additionally, in the category identification task, subjects were asked to mark the color wheel to identify the best example of each color category term (‘blue’, ‘pink’, ‘green’, ‘purple’, ‘orange’, ‘yellow’, ‘red’; see **Figure 5-1c**). The setting of this task was identical to Chapter 4, except that the color wheel consisted of 180 sample hues. Every subject finished 30 trials for each color category term. Color terms were presented in random order and the color wheel was randomly rotated in each trial.

5.2.4. A Dual-content Model of Color WM

Furthermore, we developed and implemented a probabilistic dual-content model to simulate color working memory based on a previous CATMET model from Bae and colleagues (Bae et al., 2015). It models color memorization as a combination of the visual metric pathway via von Mises distribution and the categorical pathway through probabilistic assignment based on

distribution derived from categorization tasks (**Figure 5-2**). The computation of the dual content model is explained as follows.

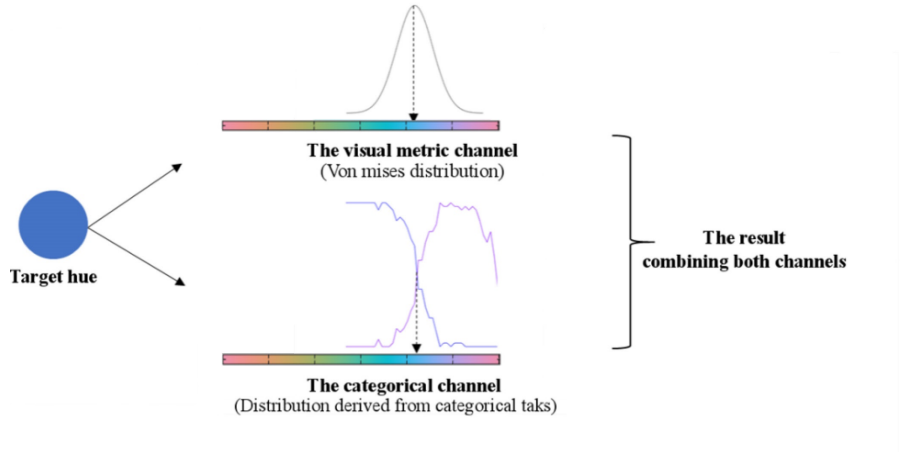

Figure 5-2 Schematic illustration of the dual-content model simulating the color working memory.

Firstly, the sample input is received as a noisy sample by the model, what is conventionally conducted in perceptual models. This is realized through applying a von Mises distribution (Φ) centered at the sample input. It can be described by

$$p(\hat{S}|S) = \Phi(\hat{S}|S + \beta, \kappa), \qquad (5\text{-}1)$$

where parameters β and κ respectively refer to the bias and the precision of the von Mises distribution. S indicates the target sample input, and $\hat{S}$ refers to the noisy sample input. $p(\hat{S}|S)$ refers to the probability of a noisy sample given the target sample. Because the noisy sample distribution centered at the target sample, the bias β equals to zero. A precision κ value of 14.89 is used for all target samples and responses based on previous evidence (Bae et al., 2015).

Secondly, the noisy sample is assigned to a category. Category boundary hues that can be best labeled with comparable probabilities by adjacent category terms are estimated in the color naming task. In this study six categories including 'pink', 'orange', 'yellow', 'green', 'blue', and 'purple' are modeled, while 'red' and 'brown' are not under consideration due to their rare usage in the category naming task (**Figure 5-4**). This results in six boundary colors between every pair of neighboring category terms. Aiming at a probabilistic category assignment, noisy borders are computed using von Mises distributions centered at the estimated borders.

$$B_i = \Phi(\mu_i, \kappa), \qquad (5\text{-}2)$$

Where μ_i refers to the empirically estimated boundary value of the ith category, κ indicates the universal precision value of 14.89. Noisy border B_i between the ith and the (i+1)th category is computed, with i varying from one to six. If i equals 6, the border B_6 between the 6th and the 1st category is calculated. With the set of noisy border values, a sample color can be assigned to a category based on its hue position relative to noisy borders. For example, the sample is assigned to ith category if its hue value lies beyond the border B_i but below the border B_{i+1}. The implementation of noisy borders and noisy samples using von Mises distributions facilitates probabilistic assignment to distinct categories on every individual simulation. To summarize the first two steps, the target sample input S is received as noisy sample $\hat{S}$ and then assigned to category $\hat{C}$.

Thirdly, the categorical output hue is estimated based on the assigned category. This is realized on the basis of a probability distribution of hues belonging to a category, which is acquired from empirical categorization tasks (Bae et al., 2015). The probability of obtaining the output hue from the categorical channel $\tilde{X}_c$ given the category $\hat{C}$ can be calculated as follows:

$$p(\tilde{X}_c|C) = \Phi(\tilde{X}_c|\mu_c, \kappa_c), \qquad (5\text{-}3)$$

Where parameters μ_c and κ_c refer to the center and the precision of the probability distribution of sample hues belonging to the category C. This distribution is obtained by combining the results from the category naming and identification tasks. The response frequency distributions about the best category terms to describe hues in the former task (**Figure 5-4**) and the frequency distributions about the hues best representing categories in the latter task (**Figure 5-5**) are converted into probability distributions and averaged. The resulting mixed probability distributions thus describe the relationship between the category terms and the color space (with 180 hues). Each of these probability distributions is fitted with a von Mises distribution. For every category, the mean μ_c and precision hue value κ_c are estimated and utilized to compute the output hue of the given category.

As next step, the visual output hue is generated through the visual metric pathway. This is computed as in typical metric models of color working memory (Zhang and Luck, 2008). A von Mises distribution is utilized to estimate the high-resolution output hue via the metric pathway. According to Bayes' rule, the probability of acquiring the output $\tilde{X}_S$ given the noisyv

input $\hat{S}$ can be estimated based on prior knowledge of related conditions (Bayes and Price, 1763):

$$p(\tilde{X}_S \,|\hat{S}) \propto p(\hat{S}|\tilde{X}_S\,)p(\tilde{X}_S). \qquad (5\text{-}4)$$

Because $p(\tilde{X}_S)$ is uniform as all color hues are presented with equivalent probabilities, $p(\tilde{X}_S \,|\hat{S})$ and $p(\hat{S}|\tilde{X}_S\,)$ are interchangeable (Bae et al., 2015). The latter can be estimated using equation (5-1).

Finally, the memorized color can be estimated from the joint probability distribution by combining the categorical pathway and the visual metric pathway. The contributing percentage of these two pathways are added to one. It is also not to forget, that trials exist where subjects respond by guessing. In case that the sample input is not encoded in memory, any hue could be chosen as the response hue. The guessing response can thus be modeled by a uniform distribution over the color space. The final response is the combined result of the memory output and the guessing part.

5.3 Results

5.3.1. Behavioral Results

Delayed Estimation Task

In the delayed estimation task, subjects were asked to mark on the color wheel consisting of 180 sample hues (with an equal spacing of 2 degrees) to best match the target hue. The response was compared with the target hue and the deviation between them is called the *bias*. The average absolute bias across all target hues and across all ten subjects is 9.28 degrees in a 360-degree circular space. A positive bias indicates that the response has a higher hue value than the target, and a negative bias indicates the opposite. For each of the 180 target hues, the response bias was estimated across all trials and all subjects (**Figure 5-3**). Our data exhibited a pattern apparently deviating from the uniform distribution. In ideal cases, if a pure visual strategy is used, the response is likely to show an approximately uniform bias pattern with random fluctuations. Thus, the color-specific bias pattern observed in this study argues against a pure visual metric representation but suggests a contributing categorization effect.

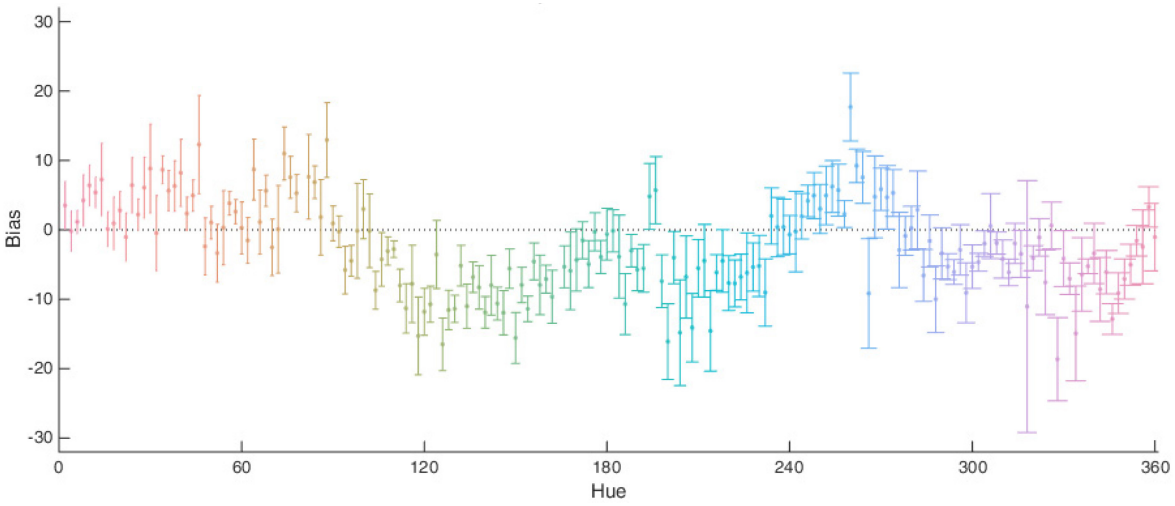

Figure 5-3 Response bias pattern in the delayed estimation task. The response bias (in degree) was estimated for every hue angle of the circular color space (in degree). Positive bias value refers to the case when the response has a higher hue value than the target hue. Error bars refer to the between-subject standard error of mean.

Category Naming Task

In the category naming task, responses regarding the best terms to describe each hue sample were collapsed across all sessions and all subjects (**Figure 5-4**). The response frequency pattern of this study showed consistency with previous studies (Boynton and Olson, 1990; Sturges and Whitfield, 1997; Bae et al., 2015) and with the fMRI sessions in Chapter 4. Based on the frequency distribution, the categorization preference in terms of boundary colors can be estimated. A hue that could be best described with equal probability by neighboring category terms is considered as a boundary hue. Two category terms – 'brown' and 'red' were rarely selected to label any color sample, and thus excluded from further analyses. Boundary hues between the remaining six color categories ('pink', 'orange', 'yellow', 'green', 'blue', 'purple') were estimated.

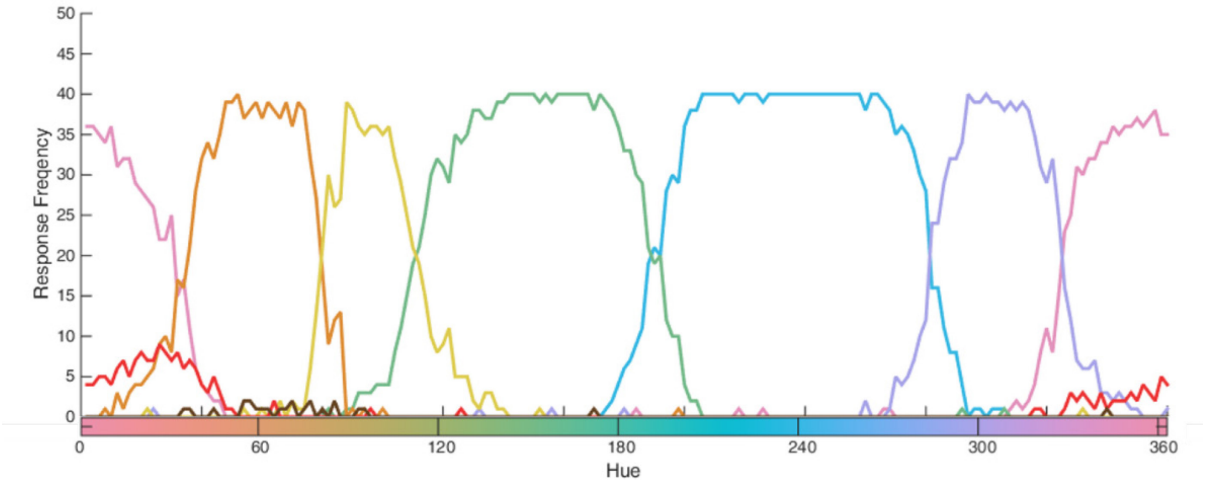

Figure 5-4 Collapsed results of the category naming task across all sessions and subjects. These response frequency distributions exhibited the frequency at which each of the eight color terms (‘pink’, ‘red’, ‘brown’, ‘orange’, ‘yellow’, ‘green’, ‘blue’, ‘purple’) was used to best describe sample hue angles of the circular color space. Boundary hue angles that can be labeled with comparable probabilities by neighboring color terms were estimated. Because the terms ‘brown’ and ‘red’ were rarely used to label any hue sample, they were excluded from further analyses.

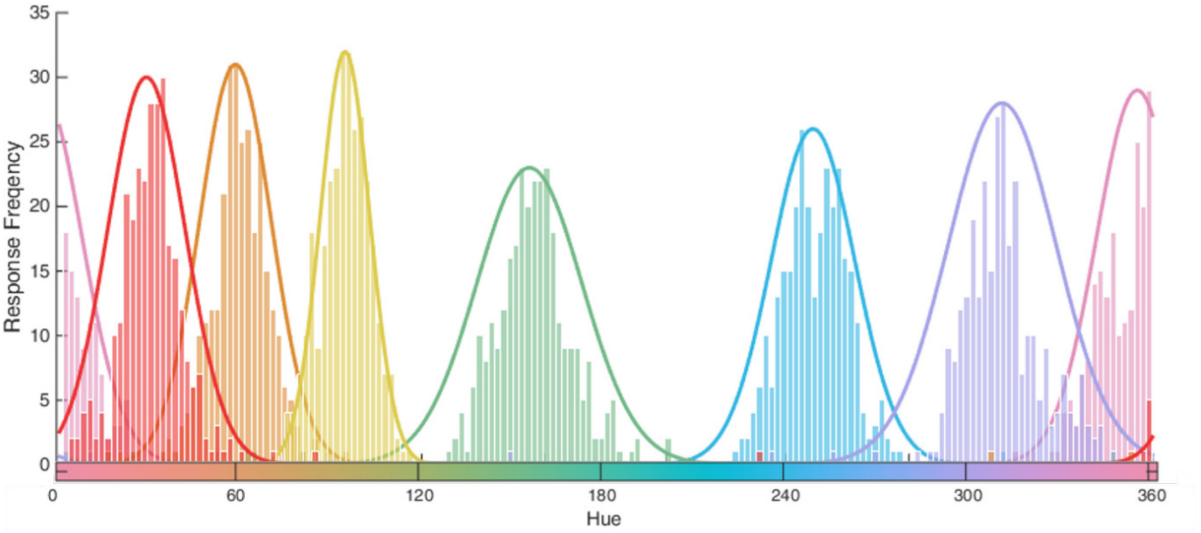

Figure 5-5 Collapsed results of the category identification task across all sessions and subjects. These response frequency distributions describe the frequency at which each hue angle from the circular color space was selected as the best exemplar of each color term (‘pink’, ‘red’, ‘orange’, ‘yellow’, ‘green’, ‘blue’, ‘purple’). Focal hues were estimated as the peak center by fitting the data with von Mises distributions (curves in respective colors).

Category Identification Task

In the category identification task, responses concerning the best hue exemplars of every color category were collapsed across all sessions and all subjects (**Figure 5-5**). The response frequency distributions showed clear peaks, which was consistent with previous evidence (Boynton and Olson, 1990; Sturges and Whitfield, 1997; Bae et al., 2015) and with Chapter 4. Compared to the uniform frequency distribution, a peak-shaped distribution indicates that some hues are predominantly selected in comparison to others to best represent a category term. The mostly frequently chosen exemplar to represent a category is called the focal hue. It is estimated by fitting the peak-shaped frequency distribution with a circular von Mises distribution. Seven terms ('pink', 'red', 'orange', 'yellow', 'green', 'blue', 'purple') were initially evaluated in this task and the term 'red' was excluded from further analyses due to its rare use in the color naming task. Because the category 'red' exhibited considerable overlap with the term 'pink' and 'orange' here, it can influence the estimation of their focal hues.

Summary of the Three Tasks

To further investigate the mechanism of color working memory, results of the three tasks were placed together (**Figure 5-6**). The bias pattern obtained in the delayed estimation task was fitted with a smoothing spline curve and examined together with the color categorization preference (focal color and boundary color) acquired from the categorization pair tasks. Ideally, if the color sample is memorized exclusively visually, the response frequency distribution should be approximately uniform over the whole circular hue space. The response distribution we acquired was clearly not uniform. Meanwhile if only a categorical strategy is used to memorize color, the response should be biased towards the focal color and biased away from the boundary color. This means that hues around the focal color exhibit low biases and hues on opposite sides of the focal color show biases of opposite directions, and that hues close to boundary color exhibit high biases (Bae et al., 2015). Our results showed an approximately systematic categorization effect, especially in some color categories. For example, the response bias can be well predicted by the relative position of the target color within the category of 'orange', 'pink' and 'yellow'. Furthermore, the bias patterns found in this study showed consistency with Bae's study (Bae et al., 2015). In summary, our results based on the three tasks provide evidence for a joint utilization of visual and categorical representations in color working memory.

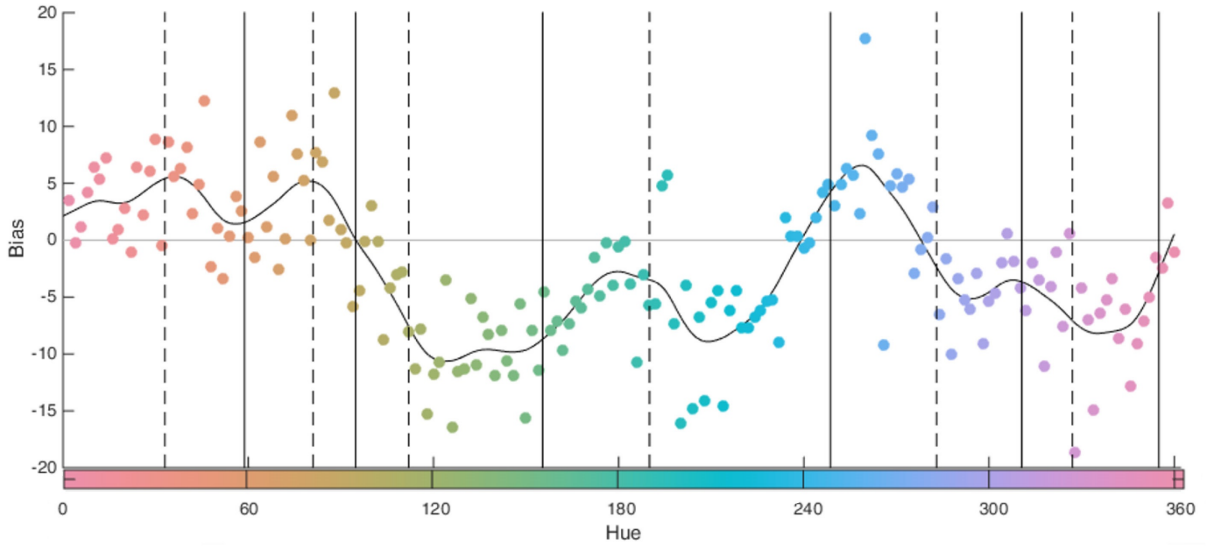

Figure 5-6 Summary of results from all three tasks. The estimates of hue-specific average bias (in degree; marked as dots in color) were superimposed with a smoothing spline curve (black curve) to illustrate the bias pattern. Vertical solid lines indicate estimated focal colors while vertical dashed lines refer to estimated boundary colors.

5.3.2. Modeling Results

By employing the dual-content model (described in 5.2.4) that combined the visual metric pathway with the categorical pathway based solely on empirical data of the color categorization tasks, the delayed response to the memory task was simulated. The guessing probability was set at 10%; and for the memorized part, the categorical and visual metric pathways were assigned with 80% and 20% probability respectively. For each of 180 sample hues, 100 simulations were conducted. The simulated bias pattern showed considerable similarity to the acquired response in the delayed estimation task (**Figure 5-7**). A significant positive correlation was found between the simulated and the empirically acquired response bias pattern ($r = 0.5379$, $P < 0.001$).

For comparison, a similar working memory model which only used the visual pathway was implemented. This model was comparably simulated 100 times for each of the sample hues. The simulated response exhibited no clear bias pattern but an approximately uniform distribution. A weak and insignificant correlation was observed in the bias pattern between the model and the experiment ($r = -0.08$, $P = 0.28$).

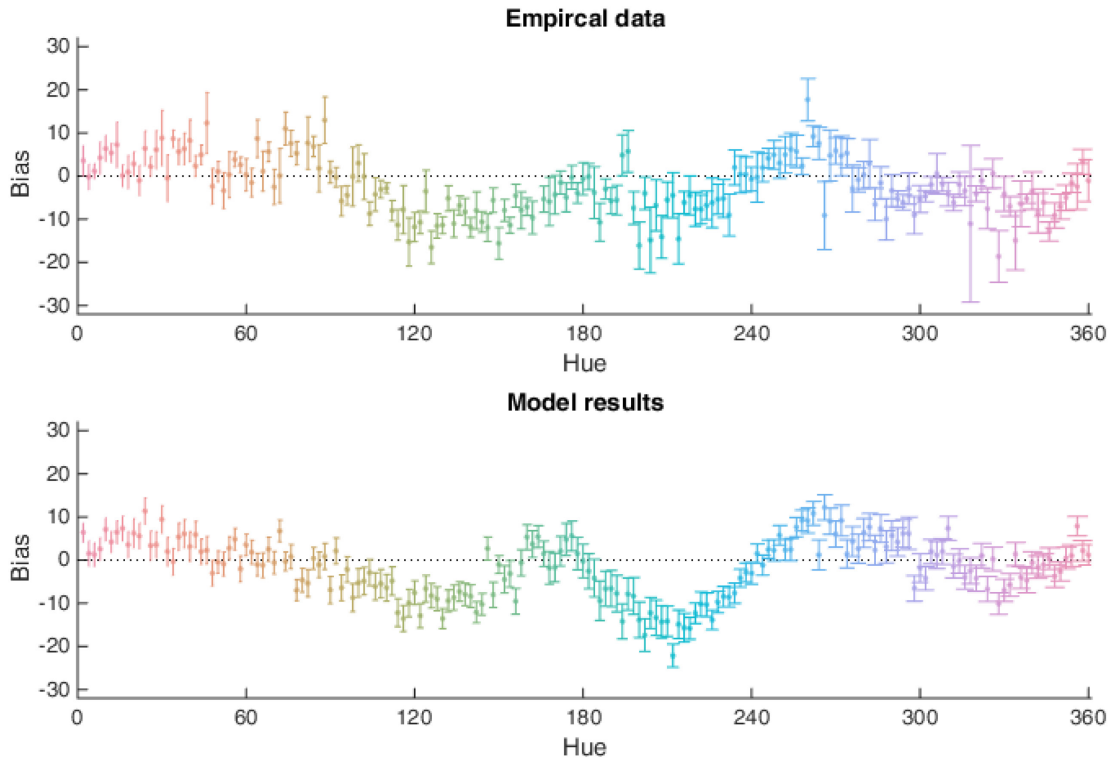

Figure 5-7 Comparison of the color-specific bias pattern (in degree) acquired from the delayed estimation experiment (top) and from the dual-content model (bottom). A significant correlation is observed between them.

5.4 Discussion

While each of the 180 equally illuminated sample hues was employed equally frequently as cued stimulus, the response exhibited a systematic stimulus-specific bias pattern. For further investigation, results from categorization tasks were put together with the delayed bias pattern. Interestingly, the bias level can be approximately predicted by the relative position of the sample hue to color categorical preferences, especially within categories of 'pink', 'orange' and 'yellow'. Around focal colors that were most frequently selected as the best exemplars of respective color categories, we discovered mostly small biases; while boundary hues which were assigned with comparable probabilities to neighboring categories often led to large biases. If only a visual strategy is used for memorization, an approximately uniform bias pattern with random fluctuation should be expected. Thus, our behavioral patterns suggest that color working memory is not limited to visual metric representation alone, and the additional contribution of categorical representation should not be ignored.

Furthermore, a dual-content model was implemented to simulate the joint representation of color working memory. This model combined the traditional visual metric pathway with the novel categorical pathway based on empirical data. The empirically acquired categorical preferences of subjects were utilized to predict the memory output in the categorical pathway. This dual-content model produced response bias patterns that showed a strong and significant correlation with experimental results. In contrast, the traditional model, which relied only on continuous visual representation, displayed a weak and insignificant correlation with experimental data. These modelling results further confirm the hypothesis of a joint dual representation in color working memory. What can be interesting to implement in the future is to predict the contributing proportion of these two strategies with this model.

The contribution of categorical representation to primary sensory representation can also exist in other feature domains. The dual-content view might be extended to orientation working memory where certain categorical labels (such as ’12 o’clock’, ‘3 o’clock’; or ‘left’, ‘right’) are memorized while at the same time the exact continuous orientation is held in memory. To test this, experimental tasks and analyses in orientation space comparable to the ones in this study can be implemented in the future. Furthermore, a dual-content probabilistic model can be constructed to investigate the combined contribution of continuous sensory representation and categorical representation in the orientation domain.

In this study, we realize that the selection of the appropriate color category terms is important but challenging. Initially, eight basic color category terms from Berlin and Kay’s, ‘blue’, ‘pink’, ‘green’, ‘purple’, ‘orange’, ‘yellow’, ‘red’, ‘brown’ (Berlin and Kay, 1969), were employed to label all 180 sample hues. However, the terms ‘brown’ and ‘red’ were so rarely used in the category naming task that they were excluded from following analyses. The exclusion of the category ‘red’, which overlapped considerably with the terms ‘pink’ and ‘orange’, could lead to an inaccurate estimation of the focal colors in the category identification task. Furthermore, subjects reported usage of a wide range of color terms. For example, they memorized the sample color by adding adjectives to the above mentioned basic terms (e.g. ‘light blue’, ‘grass-green’, ‘warm orange’), by combining terms (e.g. ‘yellow-green’), as well as by using additional color terms (e.g. ‘turquoise’, ‘watermelon’, ‘cyan’). The number of basic color terms also varied from individual to individual. To improve the model’s performance in the future, one possibility is to construct the categorical pathway by using individual-based category terms. These can be

acquired by, for example, an additional experiment, in which subjects can type freely any color term to best describe each of the 180 sample hues.

Chapter 6 General Discussion

In this final chapter, I first summarize the contents and results of the previous five chapters in brief words (section 6.1). Then, I generally discuss the three empirical studies and the advancement they bring to our understanding of working memory, as well as give a speculative outlook into the future of the research field (section 6.2).

6.1 Summary

I began this thesis by introducing the background and context of the field in Chapter 1. Human memory is comprised of three components: sensory memory, working memory and long-term memory. This chapter first described the general basics of these three kinds of human memory (section 1.1), then focused on working memory from two different angels. On the one hand, it introduced Baddeley and Hitch's cognitive model, which suggests that working memory is composed of a central executive, a visuospatial sketchpad, an episodic buffer and a phonological loop. While the central executive is the control and regulation center, the other three components are dedicated to maintaining information of the respective modalities (section 1.2). On the other hand, it addressed an ongoing debate over the neural basis of working memory maintenance and reviewed neuroimaging studies that used multivariate pattern analysis to identify candidate cortical regions for working memory storage (section 1.3).

In Chapter 2, the main methods utilized in this thesis for fMRI data acquisition and analysis were introduced. Firstly, the background and basic physical principles of fMRI were introduced (section 2.1). Secondly, the way fMRI data are preprocessed and statistically tested was briefly described (section 2.2). Thirdly, the way fMRI data are analyzed using different approaches was introduced (section 2.3). Univariate analysis (section 2.3.1) and multivariate pattern analysis (MVPA; section 2.3.2) as well as their differences were discussed (section 2.3.3). The critical method, MVPA, which identifies content-specific information from multi-voxel activity patterns was introduced in four sub-sections (sub-section 2.3.2.1 to 2.3.2.4). This chapter aimed

to provide essential methodological knowledge to precisely interpret the experimental results from fMRI studies.

In the next three chapters, I presented three empirical studies aiming to investigate content-specific working memory maintenance. Two studies employed neuroimaging and multivariate pattern analysis approaches to examine neural basis, while the other study examined behavioral patterns and constructed a probabilistic model.

In Chapter 3, an fMRI study was introduced, which examined working memory storage of a visually complex but phonetically simple language script that encouraged verbal encoding. Chinese native speakers (section 3.2.1) were required to memorize well-known Chinese characters (section 3.2.2) in a retro-cue-based match-to-sample task (section 3.2.3). Searchlight-based multivariate pattern analysis was employed to identify delayed content-specific information from multi-voxel brain activity (section 3.2.5). Broca's area and the left premotor cortex were identified to hold significant information during the delay period. These two regions were further found to carry (1) significant stimulus-specific content about cued but not uncued Chinese characters; (2) significantly more information in the left than the right hemisphere; (3) as contrast, little information about complex visual patterns that are hard to verbalize (section 3.3.3). Although the early visual cortex exhibits delayed content-specific information, it contained a comparable amount of information about cued and uncued stimuli and was thus likely to be involved in perceptual than mnemonic processes. Therefore, Broca's area and the left premotor cortex were considered as candidate stores for verbal working memory content.

Both Chapter 4 and 0 aimed to test the hypothesis that color is memorized as a combination of two sources of information: low-level visual representation and abstract categorical representation. For this purpose, a delayed estimation (DE) paradigm was employed from which the structure of color working memory can be inferred based on response biases. Furthermore, a pair of categorization tasks were employed to evaluate color categorical preferences of subjects. An additional undelayed estimation (UDE) task was utilized in Chapter 4 to obtain a clean contrast between working memory and perception processes. While sharing similar paradigms, these two chapters examined the hypothesis from distinct angles using different methods.

Chapter 4 investigated the neural basis of the dual-content representation of color. The acquired fMRI data were analyzed using multivariate pattern analysis and encoding models, which characterized selective neural response to color. This novel approach was based on the assumption that the two kinds of neural representations can be modeled by two types of encoding basis functions. The low-level visual representation can be characterized as a weighted sum of six evenly spaced conventional cosine-shaped basis functions, while the categorical representation might be modeled by a set of novel categorical basis functions. These novel functions were constructed based on empirical data from color categorization tasks. The fMRI analysis was implemented in three regions of interest (ROIs) that were found to exhibit neural color selectivity: V1, V4 and VO1. Decoding results were described in four sections.

(1) The dual-content neural representation of color was identified with two types of encoding models for respective representation (section 4.3.3). During working memory (DE task), significant low-level visual representation was estimated V1 and VO1, and significant categorical representation was found in all three ROIs. In comparison, during color perception (UDE task), significant visual representation was identified in all ROIs (consistent with Brouwer and Heeger, 2009), while categorical representation was estimated in all but VO1 region. The lack of significance was possibly but not necessarily due to the limited statistical power. In short, during both perception and working memory, color information was decoded in all three ROIs. (2) It is possible to estimate which model can decode the color representation better in every ROI (section 4.3.4). During memory (DE task), a significantly higher information measure was observed in V4 and VO1 using the empirical-based categorical encoding model. While during perception (UDE task), no significant difference was detected in any ROI. These implied a predominant categorical mnemonic representation of color in anterior regions of the visual cortex. (3) To examine whether the difference between two kinds of neural representations (decoded by two types of encoding models) in a cortical region was statistically different between delayed and undelayed estimation tasks, the interaction effect was tested with a standardized information measure (section 4.3.5). The result exhibited a significant interaction effect between two encoding models and two tasks in VO1. (4) Additionally, the empirical categorical encoding model based on color categorization preferences of respective individual subjects was compared with that based on average categorical preferences across subjects (section 4.3.6). No significant difference was observed in decoding results utilizing individual-based and average-based categorical encoding model.

0 focused on examining response bias patterns and constructing a mathematical model (Bae et al., 2015) to test the dual-content representation of color in working memory. 180 color samples (in Chapter 4 only 50 color samples due to fMRI session length limitation) with equal illuminance and even spacing in CIELAB space were utilized, resulting in covering the circular feature space in only two degree spacing (section 5.2.2). In the delayed estimation task, the subjects' responses exhibited systematic color-specific bias patterns. The bias degree of a sample color can be, to some degree, explained by its relative position compared to focal and boundary hues estimated from categorization pair tasks (section 5.3.1). A probabilistic dual-content model was constructed by utilizing empirically acquired categorical preferences to predict categorical assignment. Combining the categorical channel and the visual metric pathway, this model generated response patterns that significantly correlated with data observed in human experiments. In contrast, the conventional model based solely on continuous visual representation displayed a weak and insignificant correlation with experimental results (section 5.3.2). These results confirmed the additional contribution of categorical representation to color working memory.

6.2 General Discussion

Research on working memory has been carried out extensively in the last half century, but the nature of working memory remains widely elusive. Within the last decade, a large number of neuroimaging studies have been performed using multivariate pattern analysis (MVPA) to examine where and how working memory content is stored in the brain (Brouwer and Heeger, 2009; Harrison and Tong, 2009; Christophel et al., 2012, 2017; Jerde et al., 2012; Riggall and Postle, 2012; Emrich et al., 2013, 2013; Lee et al., 2013; Christophel and Haynes, 2014a; Ester et al., 2015; Linke and Cusack, 2015; Kumar et al., 2016). In this thesis, MVPA is also utilized as the major analysis approach, and compared to early neuroimaging studies using univariate analysis, MVPA makes it possible 1) to distinguish feature-selective from non-selective brain signals; 2) to detect content-specific cortical activity instead of contrast-specific signals (for example contrasts between English words and pseudo-words; Lewis-Peacock et al., 2012; Yue et al., 2018). Furthermore, in combination with MVPA, a set of encoding basis functions that characterized cortical feature-selectivity and enabled the decoding of a large sample number are utilized in this thesis.

While a number of studies have been performed to test content-specific working memory storage of visual (e.g. Harrison and Tong, 2009; Buschman et al., 2011; Ester et al., 2015), auditory (single tones or sounds without any semantic meanings; e.g. Linke and Cusack, 2015; Kumar et al., 2016), motion (e.g. Riggall and Postle, 2012; Emrich et al., 2013), tactile (e.g. Schmidt and Blankenburg, 2018), spatial (e.g. Jerde et al., 2012), and object information (e.g. Lee et al., 2013), direct evidence for the maintenance of verbal content has so far been missing. The first study of the thesis (Chapter 3) exhibits novel evidence that verbal working memory information of Chinese script is retained in language-related areas of Broca's area and the left premotor cortex. It is the first study to decode item-level working memory content in the verbal modality. Its findings further confirm the 'distributed' view, which argues for a coordinated recruitment of distributed region instead of centralized systems in prefrontal cortex responsible for WM storage functions (Fuster, 1995; Postle, 2006; Zimmer, 2008a; Christophel et al., 2017). Furthermore, the combined recruitment of Broca's area and the left premotor cortex might provide neuroscientific evidence to Baddeley's model, where they together serve the articulatory rehearsal process in the phonological loop (Baddeley et al., 1984).

It is the first study to decode item-level working memory content in the verbal modality. Its findings further confirm the 'distributed' view, which argues for a coordinated recruitment of distributed region instead of centralized systems in prefrontal cortex responsible for WM storage processes (Fuster, 1995; Postle, 2006; Zimmer, 2008a; Christophel et al., 2017). Furthermore, the combined recruitment of Broca's area and the left premotor cortex might provide neuroscientific evidence to Baddeley's model, where they together serve the articulatory rehearsal process in the phonological loop (Baddeley et al., 1984).

Verbal coding is not only utilized when a language script is presented, but also possible when other types of stimuli are shown. In fact, we speculate that a stimulus can be represented in various forms with different abstraction levels in the brain. For example, to memorize an object like an apple, one could retain its visual features including color, shape, size, etc., its fragrance, and its tactile feature, but one could also memorize it by the name 'apple'. It is demonstrated in the third study (0) that equally illustrated color samples are memorized with systematic bias patterns. This finding strongly suggests that color memorization is a joint combination of continuous visual representation and abstract categorical representation, consistent with previous evidence (Bae et al., 2015). Furthermore, it has been recently proposed that the storage

sites of working memory content can depend, on the one hand, on the functional roles of cortical regions, and on the other hand, on the abstraction level of the stimuli (Christophel et al., 2017).

One major goal of this thesis is to study the dissociation between low-level sensory content and abstract verbal information stored in distinct brain areas. In the second study (Chapter 4), the dual-content representation of color working memory is examined and a direct contrast between two types of representations is performed. Utilizing a conventional sensory encoding model and a novel categorical encoding model, we find significant sensory and categorical representations of color WM in V1. Comparing two types of encoding models reveals the prevalent representation form in a brain region. While anatomically moving from posterior to anterior, V1, V4, and VO1 display an elevation in the categorical representation of color working memory. Furthermore, VO1 exhibits a clear interaction effect: the prevalent categorical representation of color during working memory clearly differentiates itself from during perception. These might suggest a gradient of abstraction in the memorized content along the rostral-caudal axis of the brain. While low-level concrete sensory features are stored in posterior areas responsible for corresponding sensory perception processes, anterior areas can be alleviated from the duplication of sensory details, and thus be dedicated to retain abstract information that assist WM (Christophel et al., 2017). Previous evidence shows that prefrontal cortex is barely involved in the storage of concrete color hues, but can instead encode abstract relevant information (e.g. spatial distribution) to facilitate the precise memorization of multiple color stimuli (Lara and Wallis, 2014). This idea of labor division along the rostral-caudal axis based on abstraction level further extends the ‘distributed’ view on working memory storage (Fuster, 1995; Postle, 2006; Zimmer, 2008b; Christophel et al., 2017).

Evidence for dual-content representation also exists in other feature domains. Orientation and tactile information have been reported to be retained in language-related areas, suggesting a possible contribution of verbal coding (Ester et al., 2015; Schmidt and Blankenburg, 2018). Future studies can be conducted to test the hypothesis of dual-content neural representations of other features like orientation and tactile memory. Like in this study, one can construct the categorical encoding model to characterize categorical representation based on the empirical data of relevant naming and identification tasks, while utilizing the sensory encoding model to characterize low-level sensory representation. By integrating these two types of encoding models with multivariate pattern analysis, one can identify the information measure of both

contents. The predominant type of neural representation in a cortical region can also be evaluated by a comparison of encoding models.

Furthermore, it is possible to adjust the abstraction level of the memory content. For example, task goals (such as reporting the object category or visual details) can influence whether abstract categorical information or low-level visual details are retained in memory (Lee et al., 2013). Some visually displayed stimuli, such as Chinese characters (Chapter 3), with their complex visual appearance yet simple and well-known pronunciations, naturally encourage verbal encoding without additional instructions, if one can read them. Verbalization or abstraction of visually presented stimuli is a form of chunking that makes memorization easier (Zhang and Simon, 1985; Hue and Erickson, 1988). Although verbal coding is observed in many tasks, in some occasions, predominately sensory coding is utilized. For example, when memorizing artificial complex visual stimuli that are hard to verbalize (Cermak, 1971; Christophel et al., 2012; Christophel and Haynes, 2014b), or when the task can be easily completed with sensory coding (e.g. letter L and T; Polanía et al., 2011), one employs mainly sensory coding. It is important to be aware of the possible modality transformation, in order to precisely target the interested abstraction level by designing accordingly. A questionnaire is employed in this thesis to assist with finding the specific format of the memorized information (section 3.3.2 and 4.3.2).

Additionally, one could extend the probabilistic model that combines categorical and visual pathways (0) to evaluate the individual usage proportion of two strategies based on response patterns. Future work could be conducted utilizing this estimated strategy preference together with sensory and categorical encoding models (Chapter 4) to predict the composition of the memory content in a cortical region.

Chapter 7 References

Albers AM, Kok P, Toni I, Dijkerman HC, de Lange FP (2013) Shared representations for working memory and mental imagery in early visual cortex. Curr Biol CB 23:1427–1431.

Allefeld C, Haynes J-D (2014) Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. NeuroImage 89:345–357.

Andrews G, Halford GS (2002) A cognitive complexity metric applied to cognitive development. Cognit Psychol 45:153–219.

Anon (1964) The frontal granular cortex and behavior. New York, NY, US: McGraw-Hill.

Ashburner J, Friston KJ (2005) Unified segmentation. NeuroImage 26:839–851.

Atkinson RC, Shiffrin RM (1968) Human Memory: A Proposed System and its Control Processes. In: Psychology of Learning and Motivation (Spence KW, Spence JT, eds), pp 89–195. Academic Press. Available at: http://www.sciencedirect.com/science/article/pii/S0079742108604223.

Averbach E, Coriell AS (1961) Short-Term Memory in Vision. Bell Syst Tech J 40:309–328.

Baddeley A (1986) Working memory. New York, NY, US: Clarendon Press/Oxford University Press.

Baddeley A (1992) Working memory. Science 255:556–559.

Baddeley A (2000) The episodic buffer: a new component of working memory? Trends Cogn Sci 4:417–423.

Baddeley A (2003) Working memory: looking back and looking forward. Nat Rev Neurosci 4:829–839.

Baddeley A (2011) Working Memory: Theories, Models, and Controversies. Annu Rev Psychol 63:1–29.

Baddeley A, Lewis V, Vallar G (1984) Exploring the Articulatory Loop. Q J Exp Psychol Sect A 36:233–252.

Baddeley AD (1966) Short-term Memory for Word Sequences as a Function of Acoustic, Semantic and Formal Similarity. Q J Exp Psychol 18:362–365.

Baddeley AD, Allen RJ, Hitch GJ (2011) Binding in visual working memory: The role of the episodic buffer. Neuropsychologia 49:1393–1400.

Baddeley AD, Hitch G (1974) Working Memory. In: Psychology of Learning and Motivation (Bower GH, ed), pp 47–89. Academic Press. Available at: http://www.sciencedirect.com/science/article/pii/S0079742108604521 [Accessed February 19, 2019].

Baddeley AD, Papagno C, Vallar G (1988) When long-term learning depends on short-term storage. J Mem Lang 27:586–595.

Baddeley AD, Thomson N, Buchanan M (1975) Word length and the structure of short-term memory. J Verbal Learn Verbal Behav 14:575–589.

Bae G-Y, Olkkonen M, Allred SR, Flombaum JI (2015) Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. J Exp Psychol Gen 144:744–763.

Bae G-Y, Olkkonen M, Allred SR, Wilson C, Flombaum JI (2014) Stimulus-specific variability in color working memory with delayed estimation. J Vis 14:7–7.

Bahrick HP (1984) Semantic memory content in permastore: fifty years of memory for Spanish learned in school. J Exp Psychol Gen 113:1–29.

Bahrick HP, Bahrick PO, Wittlinger RP (1975) Fifty years of memory for names and faces: A cross-sectional approach. J Exp Psychol Gen 104:54–75.

Bandettini PA, Wong EC, Hinks RS, Tikofsky RS, Hyde JS (1992) Time course EPI of human brain function during task activation. Magn Reson Med 25:390–397.

Barrett TR, Ekstrand BR (1972) Effect of sleep on memory: III. Controlling for time-of-day effects. J Exp Psychol 96:321–327.

Bartels A, Zeki S (2000) The architecture of the colour centre in the human visual brain: new results and a review. Eur J Neurosci 12:172–193.

Barth M, Poser BA (2011) Advances in High-Field BOLD fMRI. Materials 4:1941–1955.

Bayes T, Price R (1763) LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. Philos Trans R Soc Lond 53:370–418.

Bays PM, Catalao RFG, Husain M (2009) The precision of visual working memory is set by allocation of a shared resource. J Vis 9:7.1-711.

Bays PM, Wu EY, Husain M (2011) Storage and binding of object features in visual working memory. Neuropsychologia 49:1622–1631.

Beare JI (2010) On memory and reminiscence Aristotle (ca. 350 b.c.). Ann Neurosci 17:87–91.

Berlin B, Kay P (1969) Basic Color Terms: their Universality and Evolution. Berkeley and Los Angeles: University of California Press.

Bettencourt KC, Xu Y (2016) Decoding the content of visual short-term memory under distraction in occipital and parietal areas. Nat Neurosci 19:150–157.

Bickel PJ, Freedman DA (1981) Some Asymptotic Theory for the Bootstrap. Ann Stat 9:1196–1217.

Bliss JC, Crane HD, Mansfield PK, Townsend JT (1966) Information available In brief tactile presentations. Percept Psychophys 1:273–283.

Bonferroni CE (1936) Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze. Available at: https://www.scienceopen.com/document?vid=35962296-b63d-4dac-8c75-777d5d9cc0dd [Accessed December 7, 2018].

Boynton RM, Olson CX (1990) Salience of chromatic basic color terms confirmed by three measures. Vision Res 30:1311–1317.

Brady TF, Konkle T, Gill J, Oliva A, Alvarez GA (2013) Visual long-term memory has the same limit on fidelity as visual working memory. Psychol Sci 24:981–990.

Brewer AA, Liu J, Wade AR, Wandell BA (2005) Visual field maps and stimulus selectivity in human ventral occipital cortex. Nat Neurosci 8:1102–1109.

Broca PP (1861) Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole).

Brouwer GJ, Heeger DJ (2009) Decoding and Reconstructing Color from Responses in Human Visual Cortex. J Neurosci 29:13992–14003.

Brown RM, Robertson EM (2007) Off-line processing: reciprocal interactions between declarative and procedural memories. J Neurosci Off J Soc Neurosci 27:10468–10475.

Buchsbaum BR, Olsen RK, Koch P, Berman KF (2005) Human Dorsal and Ventral Auditory Streams Subserve Rehearsal-Based and Echoic Processes during Verbal Working Memory. Neuron 48:687–697.

Buschman TJ, Siegel M, Roy JE, Miller EK (2011) Neural substrates of cognitive capacity limitations. Proc Natl Acad Sci 108:11252–11255.

Buxton RB, Wong EC, Frank LR (1998) Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn Reson Med 39:855–864.

Caplan D, Rochon E, Waters GS (1992) Articulatory and Phonological Determinants of Word Length Effects in Span Tasks. Q J Exp Psychol Sect A 45:177–192.

Carlesimo GA, Oscar-Berman M (1992) Memory deficits in Alzheimer's patients: A comprehensive review. Neuropsychol Rev 3:119–169.

Cawley GC, Talbot NLC (2010) On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. J Mach Learn Res 11:2079–2107.

Cermak GW (1971) Short-term recognition memory for complex free-form figures. Psychon Sci 25:209–211.

Christophel, Haynes (2014a) Decoding complex flow-field patterns in visual working memory. NeuroImage.

Christophel TB, Haynes J-D (2014b) Decoding complex flow-field patterns in visual working memory. NeuroImage 91:43–51.

Christophel TB, Hebart MN, Haynes J-D (2012) Decoding the Contents of Visual Short-Term Memory from Human Visual and Parietal Cortex. J Neurosci 32:12983–12989.

Christophel TB, Klink PC, Spitzer B, Roelfsema PR, Haynes J-D (2017) The Distributed Nature of Working Memory. Trends Cogn Sci 21:111–124.

Cohen J (1982) Set Correlation As A General Multivariate Data-Analytic Method. Multivar Behav Res 17:301–341.

Collins DL, Neelin P, Peters TM, Evans AC (1994) Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr 18:192–205.

Commission Internationale de l'Eclairage (1986) Colorimetry, Ed 2, CIE No. 152. Vienna: Commission Internationale de l'Eclairage.

Conrad R (1964) Acoustic Confusions in Immediate Memory. Br J Psychol 55:75–84.

Conrad R, Hull AJ (1964) Information, Acoustic Confusion and Memory Span. Br J Psychol 55:429–432.

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20:273–297.

Cosgrove KP, Mazure CM, Staley JK (2007) Evolving Knowledge of Sex Differences in Brain Structure, Function, and Chemistry. Biol Psychiatry 62:847–855.

Cowan N (2001) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Behav Brain Sci 24:87–114; discussion 114-185.

Cowan N (2012) Working Memory Capacity. Psychology Press.

Cowan N, Wood NL, Wood PK, Keller TA, Nugent LD, Keller CV (1998) Two separate verbal processing rates contributing to short-term memory span. J Exp Psychol Gen 127:141–160.

Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage 19:261–270.

Craik FIM, Lockhart RS (1972) Levels of processing: A framework for memory research. J Verbal Learn Verbal Behav 11:671–684.

Crawford LE, Huttenlocher J, Hedges LV (2006) Within-category feature correlations and Bayesian adjustment strategies. Psychon Bull Rev 13:245–250.

Darwin CJ, Turvey MT, Crowder RG (1972) An auditory analogue of the sperling partial report procedure: Evidence for brief auditory storage. Cognit Psychol 3:255–267.

Davison AC, Hinkley DV (1997) Bootstrap Methods and their Application by A. C. Davison. Camb Core Available at: /core/books/bootstrap-methods-and-their-application/ED2FD043579F27952363566DC09CBD6A [Accessed December 7, 2018].

DeYoe EA, Carman GJ, Bandettini P, Glickman S, Wieser J, Cox R, Miller D, Neitz J (1996) Mapping striate and extrastriate visual areas in human cerebral cortex. Proc Natl Acad Sci U S A 93:2382–2386.

Diekelmann S, Born J (2010) The memory function of sleep. Nat Rev Neurosci 11:114–126.

Donkin C, Kary A, Tahir F, Taylor R (2016) Resources masquerading as slots: Flexible allocation of visual working memory. Cognit Psychol 85:30–42.

Duda RO, Hart PE, Stork DG (2000) Pattern Classification. Available at: https://www.wiley.com/en-us/Pattern+Classification%2C+2nd+Edition-p-9780471056690 [Accessed December 14, 2018].

Duffau H, Capelle L, Denvil D, Gatignol P, Sichez N, Lopes M, Sichez J-P, Van Effenterre R (2003) The role of dominant premotor cortex in language: a study using intraoperative functional mapping in awake patients. NeuroImage 20:1903–1914.

Duffy S, Huttenlocher J, Hedges LV, Elizabeth Crawford L (2010) Category effects on stimulus estimation: Shifting and skewed frequency distributions. Psychon Bull Rev 17:224–230.

Dunn OJ (1961) Multiple Comparisons among Means. J Am Stat Assoc 56:52–64.

Efron B (1979) Bootstrap Methods: Another Look at the Jackknife. Ann Stat 7:1–26.

Efron B (2003) Second Thoughts on the Bootstrap. Stat Sci 18:135–140.

Ellenbogen JM, Hu PT, Payne JD, Titone D, Walker MP (2007) Human relational memory requires time and sleep. Proc Natl Acad Sci U S A 104:7723–7728.

Ellenbogen JM, Hulbert JC, Stickgold R, Dinges DF, Thompson-Schill SL (2006) Interfering with theories of sleep and memory: sleep, declarative memory, and associative interference. Curr Biol CB 16:1290–1294.

Emrich SM, Riggall AC, LaRocque JJ, Postle BR (2013) Distributed Patterns of Activity in Sensory Cortex Reflect the Precision of Multiple Items Maintained in Visual Short-Term Memory. J Neurosci 33:6516–6523.

Engel S, Zhang X, Wandell B (1997a) Colour tuning in human visual cortex measured with functional magnetic resonance imaging. Nature 388:68–71.

Engel SA, Glover GH, Wandell BA (1997b) Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb Cortex N Y N 1991 7:181–192.

Engle RW, Tuholski SW, Laughlin JE, Conway ARA (1999) Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. J Exp Psychol Gen 128:309–331.

Ester EF, Anderson DE, Serences JT, Awh E (2013) A Neural Measure of Precision in Visual Working Memory. J Cogn Neurosci 25:754–761.

Ester EF, Sprague TC, Serences JT (2015) Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. Neuron 87:893–905.

Estes WK, Taylor HA (1964) A Detection Method and Probabilistic Models for Assessing Information Processing from Brief Visual Displays. Proc Natl Acad Sci 52:446–454.

Fegen D, Buchsbaum BR, D'Esposito M (2015) The effect of rehearsal rate and memory load on verbal working memory. NeuroImage 105:120–131.

Feigenbaum EA (1961) The Simulation of Verbal Learning Behavior. Available at: https://www.rand.org/pubs/papers/P2235.html [Accessed February 8, 2019].

Fischer S, Drosopoulos S, Tsen J, Born J (2006) Implicit learning -- explicit knowing: a role for sleep in memory system interaction. J Cogn Neurosci 18:311–319.

Fischer S, Hallschmid M, Elsner AL, Born J (2002) Sleep forms memory for finger skills. Proc Natl Acad Sci U S A 99:11987–11991.

Foerde K, Poldrack RA (2009) Procedural Learning in Humans. In: Encyclopedia of Neuroscience (Squire LR, ed), pp 1083–1091. Oxford: Academic Press. Available at: http://www.sciencedirect.com/science/article/pii/B978008045046900783X [Accessed February 18, 2019].

Fougnie D, Alvarez GA (2011) Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. J Vis 11:3–3.

Fougnie D, Asplund CL, Marois R (2010) What are the units of storage in visual working memory? J Vis 10:27.

Freeman WT, Adelson EH (1991) The design and use of steerable filters. IEEE Trans Pattern Anal Mach Intell 13:891–906.

Frick RW (1988) Issues of representation and limited capacity in the visuospatial sketchpad. Br J Psychol 79:289–308.

Friston K. J., Holmes A. P., Worsley K. J., Poline J.-P., Frith C. D., Frackowiak R. S. J. (1994) Statistical parametric maps in functional imaging: A general linear approach. Hum Brain Mapp 2:189–210.

Friston KJ, Ashburner J, Frith CD, Poline J-B, Heather JD, Frackowiak RSJ (1995) Spatial registration and normalization of images. Hum Brain Mapp 3:165–189.

Friston KJ, Josephs O, Zarahn E, Holmes AP, Rouquette S, Poline J-B (2000) To Smooth or Not to Smooth?: Bias and Efficiency in fMRI Time-Series Analysis. NeuroImage 12:196–208.

Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J Neurophysiol 61:331–349.

Funahashi S, Bruce CJ, Goldman-Rakic PS (1990) Visuospatial coding in primate prefrontal neurons revealed by oculomotor paradigms. J Neurophysiol 63:814–831.

Funahashi S, Bruce CJ, Goldman-Rakic PS (1993) Dorsolateral prefrontal lesions and oculomotor delayed-response performance: evidence for mnemonic "scotomas." J Neurosci 13:1479–1497.

Fuster JM (1973) Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. J Neurophysiol 36:61–78.

Fuster JM (1995) Memory in the Cerebral Cortex: An Empirical Approach to Neural Networks in the Human and Nonhuman Primate. MIT Press 72:227–228.

Fuster JM, Alexander GE (1971) Neuron activity related to short-term memory. Science 173:652–654.

Fuster JM, Bauer RH, Jervey JP (1982) Cellular discharge in the dorsolateral prefrontal cortex of the monkey in cognitive tasks. Exp Neurol 77:679–694.

Gais S, Mölle M, Helms K, Born J (2002) Learning-dependent increases in sleep spindle density. J Neurosci Off J Soc Neurosci 22:6830–6834.

Glover GH (1999) Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage 9:416–429.

Gold J, Hahn B, Zhang W, Robinson B, Kappenman E, Beck V, Luck S (2010) Reduced capacity but spared precision and maintenance of working memory representations in schizophrenia. Arch Gen Psychiatry 67:570–577.

Goldman PS, Rosvold HE (1970) Localization of function within the dorsolateral prefrontal cortex of the rhesus monkey. Exp Neurol 27:291–304.

Goldman-Rakic PS (1995) Architecture of the Prefrontal Cortex and the Central Executive. Ann N Y Acad Sci 769:71–84.

Graf P, Schacter DL (1985) Implicit and explicit memory for new associations in normal and amnesic subjects. J Exp Psychol Learn Mem Cogn 11:501–518.

Gross CG (1963) A comparison of the effects of partial and total lateral frontal lesions on test performance by monkeys. J Comp Physiol Psychol 56:41–47.

Hadjikhani N, Liu AK, Dale AM, Cavanagh P, Tootell RBH (1998) Retinotopy and color sensitivity in human visual cortical area V8. Nat Neurosci 1:235–241.

Harrison SA, Tong (2009) Decoding reveals the contents of visual working memory in early visual areas. Nature 458:632–635.

Haxby JV (2012) Multivariate pattern analysis of fMRI: The early beginnings. NeuroImage 62:852–855.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. Science 293:2425–2430.

Haynes J-D, Rees G (2005a) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci 8:686–691.

Haynes J-D, Rees G (2005b) Predicting the Stream of Consciousness from Activity in Human Visual Cortex. Curr Biol 15:1301–1307.

Haynes J-D, Rees G (2006) Decoding mental states from brain activity in humans. Nat Rev Neurosci 7:523–534.

Hebart MN, Baker CI (2017) Deconstructing multivariate decoding for the study of brain function. NeuroImage Available at: http://www.sciencedirect.com/science/article/pii/S1053811917306523.

Hebb D (1961) Distinctive features of learning in the higher animal, In Delafresnaye, Jean Francisque. Brain mechanisms and learning. Oxford: Blackwell. pp. 37–46. Available at: https://www.scienceopen.com/document?vid=9bc61d53-e523-4832-bceb-45ba7380e6e8 [Accessed February 8, 2019].

Hebb DO (1949) The Organization of Behavior, Wiley, New York. Available at: http://voxbookra.com/the-organization-of-behavior-donald-o-hebb-books-classics-or-simple-files.pdf [Accessed February 19, 2019].

Henson R, Friston K (2016) Convolution Models for fMRI.

Henson R, Rugg DM, Friston K (2001) The Choice of Basis Functions in event-related fMRI. NeuroImage 13(6) Available at: https://www.researchgate.net/publication/2360943_The_Choice_of_Basis_Functions_in_event-related_fMRI [Accessed December 2, 2018].

Hertzog C, Dixon RA, Hultsch DF, MacDonald SWS (2003) Latent Change Models of Adult Cognition: Are Changes in Processing Speed and Working Memory Associated With Changes in Episodic Memory? Psychol Aging 18:755–769.

Heywood CA, Gadotti A, Cowey A (1992) Cortical area V4 and its role in the perception of color. J Neurosci Off J Soc Neurosci 12:4056–4065.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.

Hitch GJ (1978) The role of short-term working memory in mental arithmetic. Cognit Psychol 10:302–323.

Hooker D (1960) Plans and the structure of behavior. By George A. Miller, Eugene Galanter and Karl H. Pribram 1960. Henry Holt and company, New York. 226 pp. J Comp Neurol 115:217–217.

Hue C-W, Erickson JR (1988) Short-term memory for Chinese characters and radicals. Mem Cognit 16:196–205.

Huettel SA, Song AW, McCarthy G (2014) Functional Magnetic Resonance Imaging, Third Edition. Oxford, New York: Oxford University Press.

Hulme C, Newton P, Cowan N, Stuart G, Brown G (1999) Think before you speak: pauses, memory search, and trace redintegration processes in verbal memory span. J Exp Psychol Learn Mem Cogn 25:447–463.

Huttenlocher J, Hedges LV, Vevea JL (2000) Why do categories affect stimulus judgment? J Exp Psychol Gen 129:220–241.

Iacoboni M (2008) The role of premotor cortex in speech perception: Evidence from fMRI and rTMS. J Physiol-Paris 102:31–34.

Jacobsen CF (1935) Functions of frontal association area in primates. Arch Neurol Psychiatry 33:558–569.

Jacquemot C, Scott SK (2006) What is the relationship between phonological short-term memory and speech processing? Trends Cogn Sci 10:480–486.

Jarrold C, Bayliss DM (2007) Variation in working memory due to typical and atypical development. In: Variation in working memory, pp 134–161. New York, NY, US: Oxford University Press.

Jenkins JG, Dallenbach KM (1924) Obliviscence during Sleep and Waking. Am J Psychol 35:605–612.

Jerde TA, Merriam EP, Riggall AC, Hedges JH, Curtis CE (2012) Prioritized Maps of Space in Human Frontoparietal Cortex. J Neurosci 32:17382–17390.

Jimura K, Poldrack RA (2012) Analyses of regional-average activation and multivoxel pattern information tell complementary stories. Neuropsychologia 50:544–552.

Just MA, Carpenter PA, Keller TA, Eddy WF, Thulborn KR (1996) Brain Activation Modulated by Sentence Comprehension. Science 274:114–116.

Kamitani Y, Tong F (2005a) Decoding the visual and subjective contents of the human brain. Nat Neurosci 8:679–685.

Kamitani Y, Tong F (2005b) Decoding motion direction from activity in human visual cortex. J Vis 5:152–152.

Kinno R, Kawamura M, Shioda S, Sakai KL (2008) Neural correlates of noncanonical syntactic processing revealed by a picture-sentence matching task. Hum Brain Mapp 29:1015–1027.

Korman M, Doyon J, Doljansky J, Carrier J, Dagan Y, Karni A (2007) Daytime sleep condenses the time course of motor memory consolidation. Nat Neurosci 10:1206–1213.

Kriegeskorte N, Bandettini P (2007) Analyzing for information, not activation, to exploit high-resolution fMRI. NeuroImage 38:649–662.

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci U S A 103:3863–3868.

Kumar S, Joseph S, Gander PE, Barascud N, Halpern AR, Griffiths TD (2016) A Brain System for Auditory Working Memory. J Neurosci 36:4492–4505.

Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, Kennedy DN, Hoppel BE, Cohen MS, Turner R (1992) Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc Natl Acad Sci U S A 89:5675–5679.

Kyllonen PC, Christal RE (1990) Reasoning ability is (little more than) working-memory capacity?! Intelligence 14:389–433.

Landauer TK (1962) Rate of implicit speech. Percept Mot Skills 15:646–646.

Lara AH, Wallis JD (2014) Executive control processes underlying multi-item working memory. Nat Neurosci 17:876–883.

Lee E-Y, Cowan N, Vogel EK, Rolan T, Valle-Inclán F, Hackley SA (2010) Visual working memory deficits in patients with Parkinson's disease are due to both reduced storage capacity and impaired ability to filter out irrelevant information. Brain 133:2677–2689.

Lee S-H, Kravitz DJ, Baker CI (2013) Goal-dependent dissociation of visual and prefrontal cortices during working memory. Nat Neurosci 16:997–999.

Levy BA (1971) Role of articulation in auditory and visual short-term memory. J Verbal Learn Verbal Behav 10:123–132.

Lewis-Peacock JA, Drysdale AT, Oberauer K, Postle BR (2012) Neural Evidence for a Distinction Between Short-Term Memory and the Focus of Attention. J Cogn Neurosci 24:61–79.

Linke AC, Cusack R (2015) Flexible Information Coding in Human Auditory Cortex during Perception, Imagery, and STM of Complex Sounds. J Cogn Neurosci 27:1322–1333.

Liu T, Bai W, Yi H, Tan T, Wei J, Tian JW and X (2014) Functional Connectivity in a Rat Model of Alzheimer's Disease During a Working Memory Task. Curr Alzheimer Res Available at: http://www.eurekaselect.com/125926/article [Accessed February 19, 2019].

Logie RH, Logie RH (1995) Visuo-spatial Working Memory. Psychology Press.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature 412:150–157.

Lovatt P, Avons SE, Masterson J (2000) The Word-length Effect and Disyllabic Words. Q J Exp Psychol Sect A 53:1–22.

Luck SJ, Vogel EK (1997) The capacity of visual working memory for features and conjunctions. Nature 390:279–281.

Luck SJ, Vogel EK (2013) Visual working memory capacity: from psychophysics and neurobiology to individual differences. Trends Cogn Sci 17:391–400.

Ma WJ, Husain M, Bays PM (2014) Changing concepts of working memory. Nat Neurosci 17:347–356.

Mahalanobis PC (1936) On the Generalised Distance in Statistics. Gen Distance Stat:49–55.

Maquet P (2001) The role of sleep in learning and memory. Science 294:1048–1052.

Mardia KV, Jupp PE (2000) Directional Statistics. Wiley.

McKeefry DJ, Zeki S (1997) The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. Brain J Neurol 120 (Pt 12):2229–2242.

Meadows JC (1974) Disturbed perception of colours associated with localized cerebral lesions. Brain J Neurol 97:615–632.

Melton AW (1963) Implications of short-term memory for a general theory of memory. J Verbal Learn Verbal Behav 2:1–21.

Mikl M, Mareček R, Hluštík P, Pavlicová M, Drastich A, Chlebus P, Brázdil M, Krupa P (2008) Effects of spatial smoothing on fMRI group inferences. Magn Reson Imaging 26:490–503.

Miller GA (1956) The magical number seven plus or minus two: some limits on our capacity for processing information. Psychol Rev 63:81–97.

Ministry of Education of the People's Republic of China (1988) List of Frequently Used Characters in Modern Chinese. Available at: https://en.wikisource.org/wiki/Translation:List_of_Frequently_Used_Characters_in_Modern_Chinese [Accessed February 13, 2018].

Mitchell TM, Hutchinson R, Just MA, Niculescu RS, Pereira F, Wang X (2003) Classifying Instantaneous Cognitive States from fMRI Data. AMIA Annu Symp Proc 2003:465–469.

Mur M, Bandettini PA, Kriegeskorte N (2009) Revealing representational content with pattern-information fMRI—an introductory guide. Soc Cogn Affect Neurosci 4:101–109.

Murray DJ (1968) Articulation and acoustic confusability in short-term memory. J Exp Psychol 78:679–684.

Nairne JS (2002) Remembering Over the Short-Term: The Case Against the Standard Model. Annu Rev Psychol 53:53–81.

Niki H, Watanabe M (1976) Prefrontal unit activity and delayed response: relation to cue location versus direction of response. Brain Res 105:79–88.

Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 10:424–430.

Oberauer K, Lewandowsky S, Farrell S, Jarrold C, Greaves M (2012) Modeling working memory: An interference model of complex span. Psychon Bull Rev 19:779–819.

Oberauer K, Lin H-Y (2016) An Interference Model of Visual Working Memory. Psychol Rev.

Ogawa S, Lee TM, Kay AR, Tank DW (1990) Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Natl Acad Sci U S A 87:9868–9872.

Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, Ugurbil K (1992) Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proc Natl Acad Sci U S A 89:5951–5955.

O'Toole AJ, Jiang F, Abdi H, Haxby JV (2005) Partially Distributed Representations of Objects and Faces in Ventral Temporal Cortex. J Cogn Neurosci 17:580–590.

Park DC, Lautenschlager G, Hedden T, Davidson NS, Smith AD, Smith PK (2002) Models of visuospatial and verbal memory across the adult life span. Psychol Aging 17:299–320.

Pashler H (1988) Familiarity and visual change detection. Percept Psychophys 44:369–378.

Persuh M, Genzer B, Melara RD (2012) Iconic memory requires attention. Front Hum Neurosci 6 Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3345872/ [Accessed February 19, 2019].

Peterson LR, Johnson ST (1971) Some effects of minimizing articulation on short-term retention. J Verbal Learn Verbal Behav 10:346–354.

Petrides M, Milner B (1982) Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. Neuropsychologia 20:249–262.

Plihal W, Born J (1997) Effects of early and late nocturnal sleep on declarative and procedural memory. J Cogn Neurosci 9:534–547.

Polanía R, Paulus W, Nitsche MA (2011) Noninvasively Decoding the Contents of Visual Working Memory in the Human Prefrontal Cortex within High-gamma Oscillatory Patterns. J Cogn Neurosci 24:304–314.

Poldrack RA, Mumford JA, Nichols TE (2011) Handbook of Functional MRI Data Analysis by Russell A. Poldrack. Camb Core Available at: /core/books/handbook-of-functional-mri-data-analysis/8EDF966C65811FCCC306F7C916228529 [Accessed November 28, 2018].

Postle BR (2006) Working Memory as an Emergent Property of the Mind and Brain. Neuroscience 139:23–38.

Poudel GR, Stout JC, Domínguez D JF, Gray MA, Salmon L, Churchyard A, Chua P, Borowsky B, Egan GF, Georgiou-Karistianis N (2015) Functional changes during working memory in Huntington's disease: 30-month longitudinal data from the IMAGE-HD study. Brain Struct Funct 220:501–512.

Pratte MS, Tong F (2014) Spatial specificity of working memory representations in the early visual cortex. J Vis 14:22.

Ptito A, Crane J, Leonard G, Amsel R, Caramanos Z (1995) Visual-spatial localization by patients with frontallobe lesions invading or sparing area 46. NeuroReport 6:1781.

Pulvermüller F, Fadiga L (2010) Active perception: sensorimotor circuits as a cortical basis for language. Nat Rev Neurosci 11:351–360.

Quintana J, Yajeya J, Fuster JM (1988) Prefrontal representation of stimulus attributes during delay tasks. I. Unit activity in cross-temporal integration of sensory and sensory-motor information. Brain Res 474:211–221.

Ranganath C, Blumenfeld RS (2005) Doubts about double dissociations between short- and long-term memory. Trends Cogn Sci 9:374–380.

Riggall AC, Postle BR (2012) The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. J Neurosci Off J Soc Neurosci 32:12990–12998.

Robertson EM (2009) From Creation to Consolidation: A Novel Framework for Memory Processing. PLOS Biol 7:e1000019.

Rouder JN, Morey RD, Cowan N, Zwilling CE, Morey CC, Pratte MS (2008) An assessment of fixed-capacity models of visual working memory. Proc Natl Acad Sci U S A 105:5975–5979.

Schacter DL (1987) Implicit memory: History and current status. J Exp Psychol Learn Mem Cogn 13:501–518.

Schacter DL, Graf P (1986) Effects of elaborative processing on implicit and explicit memory for new associations. J Exp Psychol Learn Mem Cogn 12:432–444.

Schild H (1990) MRI made Easy. Schering. Available at: https://radiology.bayer.com/academy-and-training/books/mri-made-easy [Accessed November 28, 2018].

Schmidt TT, Blankenburg F (2018) Brain regions that retain the spatial layout of tactile stimuli during working memory – A 'tactospatial sketchpad'? NeuroImage 178:531–539.

Schomers MR, Kirilina E, Weigand A, Bajbouj M, Pulvermüller F (2015) Causal Influence of Articulatory Motor Cortex on Comprehending Single Spoken Words: TMS Evidence. Cereb Cortex N Y NY 25:3894–3902.

Scoville WB, Milner B (1957) LOSS OF RECENT MEMORY AFTER BILATERAL HIPPOCAMPAL LESIONS. J Neurol Neurosurg Psychiatry 20:11–21.

Serences JT, Ester EF, Vogel EK, Awh E (2009) Stimulus-Specific Delay Activity in Human Primary Visual Cortex. Psychol Sci 20:207–214.

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268:889–893.

Service E (1998) The Effect of Word Length on Immediate Serial Recall Depends on Phonological Complexity, Not Articulatory Duration. Q J Exp Psychol Sect A 51:283–304.

Shallice T, Warrington EK (1970) Independent functioning of verbal memory stores: A neuropsychological study. Q J Exp Psychol 22:261–273.

Singh K (1981) On the Asymptotic Accuracy of Efron's Bootstrap. Ann Stat 9:1187–1195.

Sneve MH, Alnæs D, Endestad T, Greenlee MW, Magnussen S (2012) Visual short-term memory: activity supporting encoding and maintenance in retinotopic visual cortex. NeuroImage 63:166–178.

Solomon SG, Lennie P (2007) The machinery of colour vision. Nat Rev Neurosci 8:276–286.

Souza AS, Rerko L, Lin H-Y, Oberauer K (2014) Focused attention improves working memory: implications for flexible-resource and discrete-capacity models. Atten Percept Psychophys 76:2080–2102.

Sperling G (1960) The information available in brief visual presentations. Psychol Monogr Gen Appl 74:1–29.

Sperling G (1963) A model for visual memory tasks. Hum Factors 5:19–31.

Spitzer B, Fleck S, Blankenburg F (2014) Parametric Alpha- and Beta-Band Signatures of Supramodal Numerosity Information in Human Working Memory. J Neurosci 34:4293–4302.

Sprague TC, Ester EF, Serences JT (2014) Reconstructions of Information in Visual Spatial Working Memory Degrade with Memory Load. Curr Biol 24:2174–2180.

Sprague TC, Ester EF, Serences JT (2016) Restoring Latent Visual Working Memory Representations in Human Cortex. Neuron 91:694–707.

Sprague TC, Serences JT (2013) Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. Nat Neurosci 16:1879–1887.

Standing L (1973) Learning 10000 pictures. Q J Exp Psychol 25:207–222.

Stickgold R, James L, Hobson JA (2000) Visual discrimination learning requires sleep after training. Nat Neurosci 3:1237–1238.

Sturges J, Whitfield TWA (1997) Salient Features of Munsell Colour Space as a Function of Monolexemic Naming and Response Latencies. Vision Res 37:307–313.

Timm NH (2002) Applied Multivariate Analysis. New York, NY: Springer.

Tulving E (1972) Episodic and semantic memory. In: Organization of memory, pp xiii, 423–xiii, 423. Oxford, England: Academic Press.

Tulving E (1989) Memory: Performance, knowledge, and experience. Eur J Cogn Psychol 1:3–26.

Tulving E, Pearlstone Z (1966) Availability versus accessibility of information in memory for words. J Verbal Learn Verbal Behav 5:381–391.

Ungerleider LG, Mishkin M (1982) Analysis of Visual Behavior (eds Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W.). MIT Press, Cambridge, Massachusetts. Available at: https://mitpress.mit.edu/books/analysis-visual-behavior [Accessed November 22, 2018].

Vallar G, Di Betta AM, Silveri MC (1997) The phonological short-term store-rehearsal system: Patterns of impairment and neural correlates. Neuropsychologia 35:795–812.

Vallar G, Papagno C (2002) The Handbook of Memory Disorders, 2nd Edition (eds Baddeley, A. D., Kopelman, M. D. & Wilson, B. A.). John Wiley & Sons Ltd. Available at: https://www.wiley.com/en-us/The+Handbook+of+Memory+Disorders%2C+2nd+Edition-p-9780471498193 [Accessed February 20, 2019].

van den Berg R, Shin H, Chou W-C, George R, Ma WJ (2012) Variability in encoding precision accounts for visual short-term memory limitations. Proc Natl Acad Sci 109:8780–8785.

Vargha-Khadem F, Gadian DG, Watkins KE, Connelly A, Paesschen WV, Mishkin M (1997) Differential Effects of Early Hippocampal Pathology on Episodic and Semantic Memory. Science 277:376–380.

Vergara J, Rivera N, Rossi-Pool R, Romo R (2016) A Neural Parametric Code for Storing Information of More than One Sensory Modality in Working Memory. Neuron 89:54–62.

Vogel EK, Woodman GF, Luck SJ (2001) Storage of features, conjunctions and objects in visual working memory. J Exp Psychol Hum Percept Perform 27:92–114.

Wagner U, Gais S, Haider H, Verleger R, Born J (2004) Sleep inspires insight. Nature 427:352–355.

Walker MP, Brakefield T, Seidman J, Morgan A, Hobson JA, Stickgold R (2003) Sleep and the time course of motor skill learning. Learn Mem Cold Spring Harb N 10:275–284.

Wandell BA, Dumoulin SO, Brewer AA (2007) Visual Field Maps in Human Cortex. Neuron 56:366–383.

Wang L, Mruczek REB, Arcaro MJ, Kastner S (2015) Probabilistic Maps of Visual Topography in Human Cortex. Cereb Cortex 25:3911–3931.

Warrington EK, Logue V, Pratt RTC (1971) The anatomical localisation of selective impairment of auditory verbal short-term memory. Neuropsychologia 9:377–387.

Warrington EK, Shallice T (1969) The selective impairment of auditory verbal short-term memory. Brain J Neurol 92:885–896.

Watanabe M (1981) Prefrontal unit activity during delayed conditional discriminations in the monkey. Brain Res 225:51–65.

Waugh NC, Norman DA (1965) Primary memory. Psychol Rev 72:89–104.

Wickelgren WA (1965) Short-term memory for phonemically similar lists. Am J Psychol 78:567–574.

Wilken P, Ma WJ (2004) A detection theory account of change detection. J Vis 4:11–11.

Willcutt EG, Doyle AE, Nigg JT, Faraone SV, Pennington BF (2005) Validity of the Executive Function Theory of Attention-Deficit/Hyperactivity Disorder: A Meta-Analytic Review. Biol Psychiatry 57:1336–1346.

Wise SP (1985) The Primate Premotor Cortex: Past, Present, and Preparatory. Annu Rev Neurosci 8:1–19.

Witzel C, Gegenfurtner KR (2013) Categorical sensitivity to color differences. J Vis 13:1–1.

Wongupparaj P, Kumari V, Morris RG (2015) The relation between a multicomponent working memory and intelligence: The roles of central executive and short-term storage functions. Intelligence 53:166–180.

Worsley KJ, Evans AC, Marrett S, Neelin P (1992) A three-dimensional statistical analysis for CBF activation studies in human brain. J Cereb Blood Flow Metab Off J Int Soc Cereb Blood Flow Metab 12:900–918.

Yue Q, Martin RC, Hamilton AC, Rose NS (2018) Non-perceptual Regions in the Left Inferior Parietal Lobe Support Phonological Short-term Memory: Evidence for a Buffer Account? Cereb Cortex Available at: https://academic.oup.com/cercor/advance-article/doi/10.1093/cercor/bhy037/4924349 [Accessed November 27, 2018].

Zeki SM (1974) Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. J Physiol 236:549–573.

Zhang G, Simon HA (1985) STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. Mem Cognit 13:193–201.

Zhang W, Luck SJ (2008) Discrete fixed-resolution representations in visual working memory. Nature 453:233–235.

Zimmer HD (2008a) Visual and spatial working memory: from boxes to networks. Neurosci Biobehav Rev 32:1373–1395.

Zimmer HD (2008b) Visual and spatial working memory: From boxes to networks. Neurosci Biobehav Rev 32:1373–1395.

Zlonoga B, Gerber A (1986) [A case from practice (49). Patient: K.F., born 6 May 1930 (bird fancier's lung)]. Schweiz Rundsch Med Prax Rev Suisse Med Prax 75:171–172.

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Berlin, 10.12.2019
Chang Yan

Statement of Authorship

I hereby declare that this dissertation is the result of my own work and that I have not used any sources other than those listed in the bibliography or specifically indicated in the text.

Berlin, December 10, 2019
Chang Yan